AFRL-IF-RS-TR-2005-238
Final Technical Report
June 2005

A SPREAD SPECTRUM APPROACH TO NEXT GENERATION INTRUSION DETECTION

State University of New York Institute of Technology

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

AIR FORCE RESEARCH LABORATORY
INFORMATION DIRECTORATE
ROME RESEARCH SITE
ROME, NEW YORK

STINFO FINAL REPORT

This report has been reviewed by the Air Force Research Laboratory, Information Directorate, Public Affairs Office (IFOIPA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

AFRL-IF-RS-TR-2005-238 has been reviewed and is approved for publication

APPROVED: /s/
JOHN FELDMAN
Project Engineer

FOR THE DIRECTOR: /s/
WARREN H. DEBANY, JR., Technical Advisor
Information Grid Division
Information Directorate

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	JUNE 2005	Final Oct 02 – Dec 03

4. TITLE AND SUBTITLE
A SPREAD SPECTRUM APPROACH TO NEXT GENERATION INTRUSION DETECTION

5. FUNDING NUMBERS
C - F30602-03-2-0023
PE - 61102F
PR - 2301
TA - 01
WU - 61

6. AUTHOR(S)
Michael J. Medley, Stella N. Batalama, and Dimitris A. Pados

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
State University of New York Institute of Technology
PO Box 3050
Utica New York 13504-3050

8. PERFORMING ORGANIZATION REPORT NUMBER
N/A

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory/IFGB
525 Brooks Road
Rome New York 13441-4505

10. SPONSORING / MONITORING AGENCY REPORT NUMBER
AFRL-IF-RS-TR-2005-238

11. SUPPLEMENTARY NOTES

AFRL Project Engineer: John Feldman/KFGB/(315) 330-2664/ John.Feldman@rl.af.mil

12a. DISTRIBUTION / AVAILABILITY STATEMENT
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

12b. DISTRIBUTION CODE

13. ABSTRACT ***(Maximum 200 Words)***

The report describes a theoretical treatment of the tradeoffs among carrying capacity, distortion, and error recovery for spread spectrum solutions to message embedding approaches. For any given host image and (block) transform domain of interest, signature vectors are derived that when used for spread-spectrum (SS) message embedding maximize the signal-to-interference-plus-noise ratio (SINR) at the output of the corresponding maximum SINR linear filter receiver. Under a (colored) Gaussian assumption on the transform domain host data, it is shown that these same signatures over minimum probability of error message recovery at any fixed host distortion level (or-conversely- minimize the host distortion for any probability of error target level) and maximize the Shannon capacity of the covert link. These results are further generalized to cover SS embedding in linearly processed transform domain host data with resulting orders of magnitude of demonstrated improvement over state of the practice SS steganographic practices. Optimized multi-signature/multi-message embedding in the same host data is addressed as well.

14. SUBJECT TERMS
Spread-Spectrum Message Embedding Capacity, Distortion, Spread Spectrum Steganography, Digital Watermarking

15. NUMBER OF PAGES
34

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

Table of **Contents**

List of Figures

1. Introduction

The term digital watermarking refers to the process of embedding a "secret" digital signal (hidden message) in another digital signal (image or audio) called "cover" or "host." Two applications of watermarking are *a)* image/audio authentication and *b)* steganography, which is an attempt to establish covert communication between trusting parties.

The first step in the design of a watermarking system is the embedding process. This is a crucial task since watermarking properties and detector design and performance depend directly on the way the watermark is inserted within the host data. While each specific watermarking application has its own requirements [1], [2], the broad objective of most steganographic applications is a satisfactory trade-off between hidden message resistance to noise/disturbance, information delivery rate, and host distortion. Message embedding can be performed either directly in the time (audio) or spatial (image) domain [3]-[6] or in a transform domain (for example, for images we may consider full-frame discrete Fourier transform (DFT) [7]-[10], full-frame discrete cosine transform (DCT) [11], block DFT or DCT [12], [13], or wavelet transforms [14]-[16]). Direct embedding in the original host signal domain may be desirable for system complexity purposes, while embedding in a transform domain may take advantage of the particular transform domain properties [17].

In this present work, we focus our attention on transform domain spread-spectrum (SS) embedding methods for image steganography. Once the transform embedding domain has been selected, the hidden message can be applied to the host data through an additive or multiplicative rule [5]-[8], [12], [18]. In the literature, additive spread-spectrum embedding methods use an amplitude modulated pseudorandom signature to deposit one information symbol across a group of host data coefficients [5], [7], [12] or a linearly transformed version of the host data coefficients [18]. In multiplicative rule SS embedding, message data multiply host data coefficients [8].

Spread-spectrum embedding algorithms for blind image steganography (that is, hidden message recovery without knowledge of the original image) have been based on the understanding that the host signal acts as a source of interference to the secret message of interest. Yet, it should also be understood that this interference is known to the message embedder. Such knowledge can be exploited appropriately to facilitate the task of the blind receiver at the other end and minimize the recovery error rate for a given host distortion level, minimize host distortion for a given target recovery error rate, maximize the Shannon capacity of the covert steganographic channel, etc. Indeed, if we were to place the steganography application in an information theoretic context, it could be viewed as a communications-

with-side-information problem [19]-[21]. Optimized embedding methods can facilitate host interference suppression at the receiver side when knowledge of the host signal is adequately exploited in system design.

In this paper, for any given image, (block) transform domain, and host bins, we derive the additive embedding signature that maximizes the output signal-to-interference-plus-noise ratio (SINR) of the linear maximum SINR receiver filter. We establish that, under a (colored) Gaussian assumption on the host bins, this same signature minimizes the receiver bit error rate (BER) at any mean square (MS) image distortion level, minimizes -conversely- the MS image distortion at any target BER, and maximizes the Shannon capacity of the covert link. We then generalize our findings to cover joint signature and linear host data projection optimization along the same lines. In this present work, we consider only scalar parameterized host data projection as in [18]. Finally, we extend signature-only as well as joint signature and host-projection optimization to multiuser (multiple signature) embedding. Our emphasis is directed primarily toward low complexity, sequential, conditional multiuser optimization.

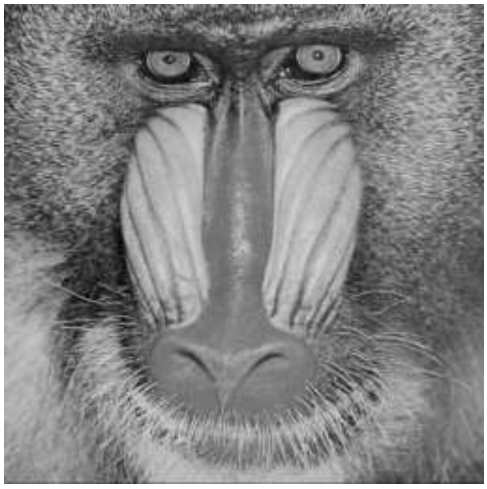
The rest of the paper is organized as follows. Section II presents our core signature and embedding optimization results. These results are generalized to multiple signature embedding in Section III. Section IV is devoted to experimental studies and comparisons. A few concluding remarks are given in Section V.

2. Signature Optimization for Spread-Spectrum Embedding

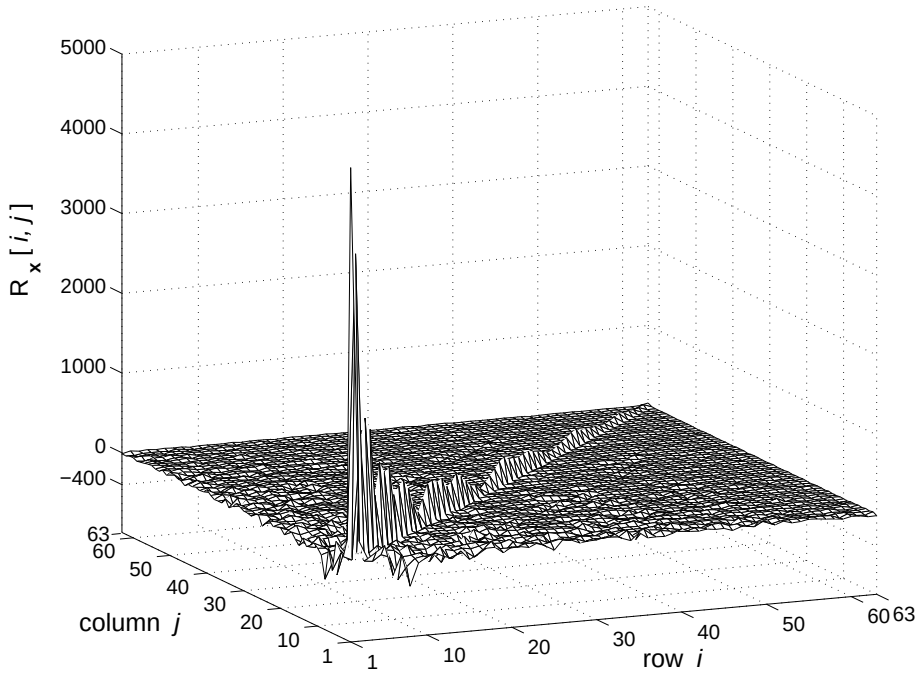
In this section we develop an optimized spread-spectrum (SS) steganographic system. To draw a parallelism with conventional SS communications systems, in SS watermark (message) embedding the watermark can be regarded as the SS signal of interest transmitted through a noisy "channel" (the host). The disturbance to the SS signal of interest is the host data themselves plus potential external noise due to physical transmission of the watermarked data and/or processing/attacking. The purpose of the watermark detector is to withstand the influence of the total end-to-end disturbance and recover the original hidden message.

A. Signal Model and Notation

Consider a host image $\mathbf{H} \in \mathcal{M}^{N_1 \times N_2}$ that is to be watermarked where $\mathcal{M}$ is the image alphabet and $N_1 \times N_2$ is the image size in pixels. Fig. 1(a) shows a grey scale baboon image example in $\mathcal{M}^{N_1 \times N_2} = \{0, 1, \cdots, 255\}^{256 \times 256}$. Without loss of generality, the image $\mathbf{H}$ is partitioned into P local blocks of size $\frac{N_1 \times N_2}{P}$ pixels. Each block $\mathbf{H}_1, \mathbf{H}_2, \cdots, \mathbf{H}_P$ is to carry one hidden information bit $b_p \in \{\pm 1\}$, $p =$



(a)



(b)

Fig. 1. (a) Baboon image example $\mathbf{H} \in \{0, 1, \cdots, 255\}^{256\times256}$. (b) Host data autocorrelation matrix (8×8 DCT, 63-bin host).

$1, 2, \cdots, P$, respectively. Embedding is performed in a real 2-dimensional transform domain $\mathcal{T}$. After transform calculation and conventional zig-zag scanning vectorization, we obtain $\mathcal{T}(\mathbf{H}_p) \in \mathbb{R}^{\frac{N_1 \times N_2}{P}}$, $p = 1, 2, \cdots, P$. From the transform domain vectors $\mathcal{T}(\mathbf{H}_p)$ we choose a fixed subset of $L \leq \frac{N_1 \times N_2}{P}$ coefficients (bins) to form the final host vectors $\mathbf{x}_p \in \mathbb{R}^L$, $p = 1, 2, \cdots, P$ (for example, it is common and appropriate to exclude the dc coefficient $\mathcal{T}(\mathbf{H}_p)[1]$ from the host $\mathbf{x}_p$).

The autocorrelation matrix of the host data $\mathbf{x}$ is an important statistical quantity for our developments and is defined as follows:

$$\mathbf{R}_x \stackrel{\triangle}{=} E\left\{\mathbf{x}\mathbf{x}^T\right\} = \frac{1}{P}\sum_{p=1}^{P} \mathbf{x}_p \mathbf{x}_p^T \tag{1}$$

where $E\{\cdot\}$ denotes statistical expectation (here, with respect to $\mathbf{x}$ over the given image $\mathbf{H}$) and T is the transpose operator. It is easy to verify that in general $\mathbf{R}_x \neq \alpha \mathbf{I}_L$, $\alpha > 0$, where $\mathbf{I}_L$ is the size-L identity matrix; that is, $\mathbf{R}_x$ *is not* constant-value diagonal or "white" in field language. For example, 8×8 DCT with 63-bin host data formation (exclude only the dc coefficient) for the baboon image in Fig. 1(a) gives the host autocorrelation matrix $\mathbf{R}_x$ in Fig. 1(b).

B. Signature Optimization

Consider direct additive SS embedding of the form

$$\mathbf{y} = Ab\mathbf{s} + \mathbf{x} + \mathbf{n} \tag{2}$$

where $A > 0$ is the bit amplitude, $\mathbf{s} \in \mathbb{R}^L$, $\|\mathbf{s}\| = 1$, is the (normalized) embedding signature to be designed, and $\mathbf{n} \sim \mathcal{N}\left(\mathbf{0}, \sigma_n^2 \mathbf{I}_L\right)$ represents potential external white Gaussian noise[1] of variance σ_n^2. The mean squared distortion of the original image *due to the watermark only* is

$$\mathcal{D} = E\left\{\|Ab\mathbf{s} + \mathbf{x} - \mathbf{x}\|^2\right\} = A^2. \tag{3}$$

With signal of interest $Ab\mathbf{s}$ and total disturbance $\mathbf{x} + \mathbf{n}$ in (2), the linear filter that operates on $\mathbf{y}$ and offers maximum SINR at its output is

$$\mathbf{w}_{\text{maxSINR}} = \arg\max_{\mathbf{w}} \frac{E\left\{\left\|\mathbf{w}^T(Ab\mathbf{s})\right\|^2\right\}}{E\left\{\left\|\mathbf{w}^T(\mathbf{x}+\mathbf{n})\right\|^2\right\}} = \left(\mathbf{R}_x + \sigma_n^2 \mathbf{I}_L\right)^{-1} \mathbf{s}. \tag{4}$$

The exact maximum output SINR value attained is

$$\text{SINR}_{\max} = \frac{E\left\{\left\|\mathbf{s}^T(\mathbf{R}_x + \sigma_n^2 \mathbf{I}_L)^{-1} Ab\mathbf{s}\right\|^2\right\}}{E\left\{\left\|\mathbf{s}^T(\mathbf{R}_x + \sigma_n^2 \mathbf{I}_L)^{-1}(\mathbf{x}+\mathbf{n})\right\|^2\right\}} = A^2 \mathbf{s}^T\left(\mathbf{R}_x + \sigma_n^2 \mathbf{I}_L\right)^{-1}\mathbf{s}. \tag{5}$$

[1]Additive white Gaussian noise is frequently viewed as a suitable model for quantization errors, channel transmission disturbances, and/or image processing attacks.

We propose to view $\text{SINR}_{\max}$ as a function of the embedding signature $\mathbf{s}$, $\text{SINR}_{\max}(\mathbf{s})$, and identify the signature that maximizes the SINR at the output of the maximum SINR filter. Our findings are presented in the form of a proposition below. The proof is straightforward and, therefore, omitted.

Proposition 1 *Consider additive SS embedding according to (2). Let* $\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_L$ *be eigenvectors of* $\mathbf{R}_x$ *in (1) with corresponding eigenvalues* $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_L$*. For any watermark induced distortion level* $\mathcal{D}$*, a signature that maximizes the output SINR of the maximum SINR filter is*

$$\mathbf{s}^{opt} = \arg\max_{\mathbf{s}} \left\{ A^2 \mathbf{s}^T \left(\mathbf{R}_x + \sigma_n^2 \mathbf{I}_L \right)^{-1} \mathbf{s} \right\} = \mathbf{q}_L. \tag{6}$$

When $\mathbf{s} = \mathbf{q}_L$*, the output SINR is maximized to*

$$SINR_{max}(\mathbf{q}_L) = A^2 \mathbf{q}_L^T (\mathbf{R}_x + \sigma_n^2 \mathbf{I}_L)^{-1} \mathbf{q}_L = \frac{A^2}{\lambda_L + \sigma_n^2} = \frac{\mathcal{D}}{\lambda_L + \sigma_n^2} \tag{7}$$

and maximum SINR data filtering simplifies to

$$\mathbf{w}_{maxSINR}^T \mathbf{y} = \mathbf{q}_L^T (\mathbf{R}_x + \sigma_n^2 \mathbf{I}_L)^{-1} \mathbf{y} \equiv \mathbf{q}_L^T \mathbf{y}. \quad \square \tag{8}$$

In summary, Proposition 1 says that the "minimum" eigenvector of the host data autocorrelation matrix, when used as the embedding signature, sends the output SINR to its maximum possible value $\frac{\mathcal{D}}{\lambda_L + \sigma_n^2}$. At the same time, maximum SINR filtering becomes plain signature (eigenvector) matched filtering.

If, in addition, we are allowed to assume that $\mathbf{x}$ is Gaussian, $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_x)$, then

$$\hat{b} = \text{sign}(\mathbf{w}_{\text{maxSINR}}^T \mathbf{y}) \tag{9}$$

is the optimum (minimum probability of error) bit detector [22] with probability of error

$$P_e = Q\left(A \sqrt{\mathbf{s}^T (\mathbf{R}_x + \sigma_n^2 \mathbf{I}_L)^{-1} \mathbf{s}} \right) = Q\left(\sqrt{\text{SINR}_{\max}(\mathbf{s})} \right) \tag{10}$$

where $Q(a) = \int_a^\infty \frac{1}{\sqrt{2\pi}} \exp^{\frac{-\tau^2}{2}} d\tau$. We see that P_e is a monotonically decreasing function of $\text{SINR}_{\max}$. If we now view P_e as a function of the embedding signature $\mathbf{s}$, $P_e(\mathbf{s})$, then Proposition 2 below follows directly from Proposition 1 and (10).

Proposition 2 *Consider additive SS embedding according to (2) with* $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_x)$*. Let* $\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_L$ *be eigenvectors of* $\mathbf{R}_x$ *with corresponding eigenvalues* $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_L$*. For any watermark induced distortion level* $\mathcal{D}$*, a signature that minimizes the probability of error of the optimum bit detector is*

$$\mathbf{s}^{opt} = \arg\min_{\mathbf{s}} \left\{ Q\Big(A \sqrt{\mathbf{s}^T (\mathbf{R}_x + \sigma_n^2 \mathbf{I}_L)^{-1} \mathbf{s}} \Big) \right\} = \mathbf{q}_L. \tag{11}$$

When $\mathbf{s} = \mathbf{q}_L$*, the probability of error of the optimum detector is minimized to*

$$P_e(\mathbf{q}_L) = Q\left(\sqrt{\frac{\mathcal{D}}{\lambda_L + \sigma_n^2}}\right) \tag{12}$$

and optimum detection reduces to

$$\hat{b} = sign\left(\mathbf{q}_L^T \mathbf{y}\right). \tag{13}$$

Conversely, for any preset probability of error level P_e*,* $\mathbf{s} = \mathbf{q}_L$ *minimizes the watermark induced distortion to*

$$\mathcal{D} = \left(\lambda_L + \sigma_n^2\right)\left[Q^{-1}\left(P_e\right)\right]^2. \quad \square \tag{14}$$

Proposition 2 explains that under a Gaussian host data assumption the "minimum" eigenvector of the host data autocorrelation matrix, when used as the embedding signature, allows message recovery with the minimum possible bit error rate $Q\left(\sqrt{\frac{\mathcal{D}}{\lambda_L+\sigma_n^2}}\right)$ and trivial signature (eigenvector) matched filter detection. Conversely, the watermark induced image distortion $\mathcal{D}$ is minimized for any given target bit error rate.

If necessary, further bit error rate improvements below $Q\left(\sqrt{\frac{\mathcal{D}}{\lambda_L+\sigma_n^2}}\right)$ for any fixed distortion $\mathcal{D}$ can be attained via error correcting coding of the information bits at the expense of reduced information bit payload. The maximum possible payload in bits that still allows -theoretically for asymptotically large number of image blocks P- message recovery with arbitrarily small probability of error is CP where $C = \max_{f_b} I(b; \mathbf{y})$ is the Shannon capacity of the covert link in bits per embedding attempt. We recall that $I(b; \mathbf{y})$ identifies the information conveyed about the embedded bit b by the received vector $\mathbf{y}$ and f_b denotes the bit probability distribution function. For Gaussian host data $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_x)$ and an average image distortion constraint $\mathcal{D}$, we can calculate [23]

$$C = \frac{1}{2}\log\det\left(\mathbf{I}_L + \mathcal{D}\left(\mathbf{R}_x + \sigma_n^2\mathbf{I}_L\right)^{-1}\mathbf{s}\mathbf{s}^T\right) \tag{15}$$

where $\det(\cdot)$ is the determinant operator. We can show that the signature choice $\mathbf{s} = \mathbf{q}_L$ is also optimal in terms of capacity, i.e. maximizes the capacity C of the covert link and therefore the maximum allowable payload CP for the host vectors $\mathbf{x}_p \in \mathbb{R}^L$, $p = 1, 2, \cdots, P$. The result is presented in the form of Proposition 3 below whose proof is given in the Appendix.

Proposition 3 *Consider additive SS embedding according to (2) with* $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_x)$*. Let* $\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_L$ *be eigenvectors of* $\mathbf{R}_x$ *with corresponding eigenvalues* $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_L$*. For any watermark induced distortion level* $\mathcal{D}$*, a signature that maximizes the covert channel capacity is*

$$\mathbf{s}^{opt} = \arg\max_{\mathbf{s}}\left\{\frac{1}{2}\log\det\left(\mathbf{I}_L + \mathcal{D}\left(\mathbf{R}_x + \sigma_n^2\mathbf{I}_L\right)^{-1}\mathbf{s}\mathbf{s}^T\right)\right\} = \mathbf{q}_L. \tag{16}$$

When $\mathbf{s} = \mathbf{q}_L$*, the covert channel capacity is maximized to*

$$C(\mathbf{q}_L) = \frac{1}{2} log\left(1 + \frac{\mathcal{D}}{\lambda_L + \sigma_n^2}\right) \quad \textit{bits per bit embedding.} \quad \square \tag{17}$$

C. Signature Optimization for Linearly Transformed Host Data

In this section we generalize the previous developments and assume that the host data vector $\mathbf{x}$ can be linearly transformed by an $L \times L$ operator of the form $\mathbf{I}_L - c\mathbf{s}\mathbf{s}^T$ [18] where both the parameter $c \in \mathbb{R}$ and the signature $\mathbf{s} \in \mathbb{R}^L$, $\|\mathbf{s}\| = 1$, are to be designed[2]. In parallel to (2), the composite signal now becomes

$$\mathbf{y} = Ab\mathbf{s} + (\mathbf{I}_L - c\mathbf{s}\mathbf{s}^T)\mathbf{x} + \mathbf{n} \tag{18}$$

and the mean squared distortion *due to the watermarking operation only* is

$$\mathcal{D} = E\left\{\left\|Ab\mathbf{s} + \left(\mathbf{I}_L - c\mathbf{s}\mathbf{s}^T\right)\mathbf{x} - \mathbf{x}\right\|^2\right\} = E\left\{\left\|\left(Ab - c\mathbf{s}^T\mathbf{x}\right)\mathbf{s}\right\|^2\right\} = A^2 + c^2\mathbf{s}^T\mathbf{R}_x\mathbf{s}. \tag{19}$$

We observe that, in contrast to (3), the distortion level is controlled not only by A but by $\mathbf{s}$ and c as well.

With signal of interest $Ab\mathbf{s}$ and total disturbance $(\mathbf{I}_L - c\mathbf{s}\mathbf{s}^T)\mathbf{x} + \mathbf{n}$ in (18), the linear filter that operates on $\mathbf{y}$ and offers maximum SINR at its output is

$$\mathbf{w}_{\text{maxSINR}} = \arg\max_{\mathbf{w}} \frac{E\left\{\left\|\mathbf{w}^T(Ab\mathbf{s})\right\|^2\right\}}{E\left\{\left\|\mathbf{w}^T((\mathbf{I}_L - c\mathbf{s}\mathbf{s}^T)\mathbf{x} + \mathbf{n})\right\|^2\right\}} = \left(\left(\mathbf{I}_L - c\mathbf{s}\mathbf{s}^T\right)\mathbf{R}_x\left(\mathbf{I}_L - c\mathbf{s}\mathbf{s}^T\right) + \sigma_n^2\mathbf{I}_L\right)^{-1}\mathbf{s}. \tag{20}$$

The exact maximum output SINR value attained is

$$\text{SINR}_{\max} = A^2\mathbf{s}^T\left(\left(\mathbf{I}_L - c\mathbf{s}\mathbf{s}^T\right)\mathbf{R}_x\left(\mathbf{I}_L - c\mathbf{s}\mathbf{s}^T\right) + \sigma_n^2\mathbf{I}_L\right)^{-1}\mathbf{s}. \tag{21}$$

In the following, we look at $\text{SINR}_{\max}$ as a function of both the embedding signature $\mathbf{s}$ and the parameter c, $\text{SINR}_{\max}(\mathbf{s}, c)$, and identify the signature and parameter values that jointly maximize the SINR at the output of the maximum SINR filter. Our findings are presented in the form of a proposition below. The proof is given in the Appendix.

Proposition 4 *Consider additive SS embedding in linearly transformed host data according to (18). Let* $\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_L$ *be eigenvectors of* $\mathbf{R}_x$ *with corresponding eigenvalues* $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_L$*. For any watermark induced distortion level* $\mathcal{D}$*, a (signature* $\mathbf{s}$*, parameter* c*) pair that maximizes the output SINR of the maximum SINR filter is*

$$\mathbf{s}^{opt} = \arg\max_{\mathbf{s}}\left\{A^2\mathbf{s}^T\left(\left(\mathbf{I}_L - c\mathbf{s}\mathbf{s}^T\right)\mathbf{R}_x\left(\mathbf{I}_L - c\mathbf{s}\mathbf{s}^T\right) + \sigma_n^2\mathbf{I}_L\right)^{-1}\mathbf{s}\right\} = \mathbf{q}_L \tag{22}$$

[2]If $c = 0$, we revert to the developments of Section II.B. If $c = 1$, $\mathbf{I}_L - c\mathbf{s}\mathbf{s}^T$ becomes the projector orthogonal to $\mathbf{s}$.

and

$$c^{opt} = \frac{\lambda_L + \sigma_n^2 + \mathcal{D} - \sqrt{\left(\lambda_L + \sigma_n^2 + \mathcal{D}\right)^2 - 4\lambda_L \mathcal{D}}}{2\lambda_L}. \tag{23}$$

When $\mathbf{s} = \mathbf{q}_L$ *and* $c = c^{opt}$*, the output SINR is maximized to*

$$SINR_{max}(\mathbf{q}_L, c^{opt}) = \frac{\mathcal{D} - c^{opt2}\lambda_L}{\lambda_L \left(1 - c^{opt}\right)^2 + \sigma_n^2} \tag{24}$$

and maximum SINR data filtering simplifies to

$$\mathbf{w}_{maxSINR}^T \mathbf{y} = \mathbf{q}_L^T \left(\left(\mathbf{I}_L - c^{opt}\mathbf{q}_L\mathbf{q}_L^T\right)\mathbf{R}_x \left(\mathbf{I}_L - c^{opt}\mathbf{q}_L\mathbf{q}_L^T\right) + \sigma_n^2\mathbf{I}_L\right)^{-1}\mathbf{y} \equiv \mathbf{q}_L^T\mathbf{y}. \tag{25}$$

The target distortion $\mathcal{D}$ *is achieved when the bit amplitude is set at* $A = \sqrt{\mathcal{D} - c^{opt2}\lambda_L}$*.* □

Proposition 4 shows that the optimum signature assignment is still the "minimum" eigenvector of $\mathbf{R}_x$ and maximum SINR filtering still reduces to plain signature (eigenvector) matched filtering. The optimum selection of c depends on the minimum eigenvalue of $\mathbf{R}_x$, λ_L, the noise variance σ_n^2, and the target distortion level $\mathcal{D}$. The optimum pair $(\mathbf{s}^{\text{opt}}, c^{\text{opt}})$ allows the output SINR to attain its maximum possible value $\frac{\mathcal{D} - c^{\text{opt}2}}{\lambda_L(1-c^{\text{opt}})^2 + \sigma_n^2}$.

If we assume that $\mathbf{x}$ is Gaussian, $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_x)$, then $\hat{b} = \text{sign}\left(\mathbf{w}_{\text{maxSINR}}^T \mathbf{y}\right)$ is the optimum bit detector with probability of error

$$P_e = Q\left(A\sqrt{\mathbf{s}^T\left(\left(\mathbf{I}_L - c\mathbf{s}\mathbf{s}^T\right)\mathbf{R}_x\left(\mathbf{I}_L - c\mathbf{s}\mathbf{s}^T\right) + \sigma_n^2\mathbf{I}_L\right)^{-1}\mathbf{s}}\right) = Q\left(\sqrt{\text{SINR}_{\max}(\mathbf{s}, c)}\right). \tag{26}$$

As in the plain additive SS embedding scenario, if $\mathbf{x}$ is Gaussian then the probability of error is a monotonically decreasing function of $\text{SINR}_{\max}$. The pair $(\mathbf{s}^{\text{opt}}, c^{\text{opt}})$ which maximizes the output SINR of the maximum SINR filter is also minimizing the probability of error of the optimum detector. The details are given in the following proposition.

Proposition 5 *Consider additive SS embedding according to (18) with* $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_x)$*. Let* $\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_L$ *be eigenvectors of* $\mathbf{R}_x$ *with corresponding eigenvalues* $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_L$*. For any watermark induced distortion level* $\mathcal{D}$*, a (signature* $\mathbf{s}$*, parameter* c*) pair that minimizes the probability of error of the optimum bit detector is*

$$\mathbf{s}^{opt} = \arg\min_{\mathbf{s}} \left\{ Q\left(A\sqrt{\mathbf{s}^T\left(\left(\mathbf{I}_L - c\mathbf{s}\mathbf{s}^T\right)\mathbf{R}_x\left(\mathbf{I}_L - c\mathbf{s}\mathbf{s}^T\right) + \sigma_n^2\mathbf{I}_L\right)^{-1}\mathbf{s}}\right)\right\} = \mathbf{q}_L \tag{27}$$

and

$$c^{opt} = \frac{\lambda_L + \sigma_n^2 + \mathcal{D} - \sqrt{\left(\lambda_L + \sigma_n^2 + \mathcal{D}\right)^2 - 4\lambda_L \mathcal{D}}}{2\lambda_L}. \tag{28}$$

When $\mathbf{s} = \mathbf{q}_L$ *and* $c = c^{opt}$*, the probability of error of the optimum detector is minimized to*

$$P_e(\mathbf{q}_L, c^{opt}) = Q\left(\sqrt{\frac{\mathcal{D} - c^{opt2}\lambda_L}{\lambda_L\left(1 - c^{opt}\right)^2 + \sigma_n^2}}\right) \tag{29}$$

and optimum detection reduces to $\hat{b} = sign(\mathbf{q}_L^T\mathbf{y})$*.* □

Proposition 5 implies that the "minimum" eigenvector of the host data autocorrelation matrix when used as the embedding signature together with c^{opt} allows message recovery with the minimum possible probability of error $Q\left(\sqrt{\frac{\mathcal{D}-c^{\text{opt}2}\lambda_L}{\lambda_L(1-c^{\text{opt}})^2+\sigma_n^2}}\right)$ (conversely, the induced distortion $\mathcal{D}$ is minimized for a given target probability of error P_e). We can show that for Gaussian host data $\mathbf{x}$, the covert channel capacity is given by

$$C = \frac{1}{2}\log\det\left(\mathbf{I}_L + \left(\mathcal{D} - c^2\mathbf{s}^T\mathbf{R}_x\mathbf{s}\right)\left(\left(\mathbf{I}_L - c\mathbf{s}\mathbf{s}^T\right)\mathbf{R}_x\left(\mathbf{I}_L - c\mathbf{s}\mathbf{s}\right) + \sigma_n^2\mathbf{I}_L\right)^{-1}\mathbf{s}\mathbf{s}^T\right). \tag{30}$$

Then, we can prove that the $(\mathbf{s}^{\text{opt}}, c^{\text{opt}})$ assignment of Proposition 5 is also optimal in terms of capacity. This result is summarized in the following proposition whose proof is given in the Appendix.

Proposition 6 *Consider additive SS embedding according to (18) with* $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_x)$*. Let* $\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_L$ *be eigenvectors of* $\mathbf{R}_x$ *with corresponding eigenvalues* $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_L$*. For any watermark induced distortion level* $\mathcal{D}$*, a (signature* $\mathbf{s}$*, parameter* c*) pair that maximizes the covert channel capacity is*

$$\mathbf{s}^{opt} = \arg\max_{\mathbf{s}}\left\{\mathbf{I}_L + \left(\mathcal{D} - c^2\mathbf{s}^T\mathbf{R}_x\mathbf{s}\right)\left(\left(\mathbf{I}_L - c\mathbf{s}\mathbf{s}^T\right)\mathbf{R}_x\left(\mathbf{I}_L - c\mathbf{s}\mathbf{s}\right) + \sigma_n^2\mathbf{I}_L\right)^{-1}\mathbf{s}\mathbf{s}^T\right\} = \mathbf{q}_L \tag{31}$$

and

$$c^{opt} = \frac{\lambda_L + \sigma_n^2 + \mathcal{D} - \sqrt{\left(\lambda_L + \sigma_n^2 + \mathcal{D}\right)^2 - 4\lambda_L\mathcal{D}}}{2\lambda_L}. \tag{32}$$

When $\mathbf{s} = \mathbf{q}_L$ *and* $c = c^{opt}$*, the covert channel capacity is maximized to*

$$C(\mathbf{q}_L, c^{opt}) = \frac{1}{2}\,log\left(1 + \frac{\mathcal{D} - c^{opt2}\lambda_L}{\lambda_L(1 - c^{opt})^2 + \sigma_n^2}\right) \quad bits\ per\ bit\ embedding. \quad \square \tag{33}$$

3. Multi-Signature Embedding

We may generalize the signal model in (2) to cover multi-signature/multi-message embedding of the form

$$\mathbf{y} = \sum_{i=1}^{K} A_i b_i \mathbf{s}_i + \mathbf{x} + \mathbf{n} \tag{34}$$

where bits $b_1, b_2, \ldots, b_K$, coming potentially from K distinct messages, are embedded simultaneously in $\mathbf{x}$ with corresponding amplitudes $A_i > 0$ and embedding signatures $\mathbf{s}_i \in \mathbb{R}^L$, $\|\mathbf{s}_i\| = 1$, $i = 1, 2, \ldots, K$. Thus, the contribution of each individual embedded message bit b_i to the composite watermarked signal is $A_i b_i \mathbf{s}_i$ and the mean squared distortion to the original host data $\mathbf{x}$ due to the embedded message i alone is

$$\mathcal{D}_i = E\left\{\|A_i b_i \mathbf{s}_i\|^2\right\} = A_i^2, \;\; i = 1, 2, \ldots, K. \tag{35}$$

Under a statistical independence assumption across message bits, the mean squared distortion of the original image due to the total multi-message watermark is

$$\mathcal{D} = E\left\{\left\|\sum_{i=1}^{K} A_i b_i \mathbf{s}_i\right\|^2\right\} = \sum_{i=1}^{K} {A_i}^2. \tag{36}$$

With signal of interest $A_j b_j \mathbf{s}_j$ and total disturbance $\sum_{\substack{i=1 \\ i\neq j}}^{K} A_i b_i \mathbf{s}_i + \mathbf{x} + \mathbf{n}$ in (34), the linear filter that operates on $\mathbf{y}$ and offers maximum SINR at its output is

$$\mathbf{w}_{\text{maxSINR},j} = \mathbf{R}_{/j}^{-1} \mathbf{s}_j \tag{37}$$

where $\mathbf{R}_{/j}$ is the "exclude-j" data autocorrelation matrix, that is the autocorrelation matrix of the disturbance to message j defined as

$$\mathbf{R}_{/j} \stackrel{\triangle}{=} E\left\{\left(\sum_{\substack{i=1 \\ i\neq j}}^{K} A_i b_i \mathbf{s}_i + \mathbf{x} + \mathbf{n}\right)\left(\sum_{\substack{i=1 \\ i\neq j}}^{K} A_i b_i \mathbf{s}_i + \mathbf{x} + \mathbf{n}\right)^T\right\} = \sum_{\substack{i=1 \\ i\neq j}}^{K} {A_i}^2 \mathbf{s}_i {\mathbf{s}_i}^T + \mathbf{R}_x + \sigma_n^2 \mathbf{I}_L. \tag{38}$$

The exact maximum output SINR value attained is

$$\text{SINR}_{\max,j} = A_j^2 \mathbf{s}_j^T \mathbf{R}_{/j}^{-1} \mathbf{s}_j. \tag{39}$$

As in Section II for single-message embedding, we propose to view $\text{SINR}_{\max,j}$ as a function of the embedding signature $\mathbf{s}_j$, $\text{SINR}_{\max,j}(\mathbf{s}_j)$, and identify the signature vector that maximizes the SINR value. Our findings are presented in the form of Proposition 7 below that parallels the developments of Proposition 1 for the single-message case.

Proposition 7 *Consider additive SS embedding according to (34). Let $\mathbf{q}_1, \mathbf{q}_2, \cdots \mathbf{q}_L$ be eigenvectors of $\mathbf{R}_{/j}$ in (38) with corresponding eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_L$. For any message induced distortion level $\mathcal{D}_j$, a signature that maximizes the output SINR of the maximum SINR filter $\mathbf{w}_{maxSINR,j}$ is*

$$\mathbf{s}_j^{opt} = \arg\max_{\mathbf{s}} \left\{A_j^2 \mathbf{s}^T \mathbf{R}_{/j}^{-1} \mathbf{s}\right\} = \mathbf{q}_L. \tag{40}$$

When $\mathbf{s}_j = \mathbf{q}_L$*, the output SINR is maximized to*

$$SINR_{max,j}(\mathbf{q}_L) = \frac{A_j^2}{\lambda_L + \sigma_n^2} = \frac{\mathcal{D}_j}{\lambda_L + \sigma_n^2} \tag{41}$$

and maximum SINR data filtering simplifies to $\mathbf{w}_{maxSINR,j}^T \mathbf{y} = \mathbf{q}_L^T \mathbf{R}_{/j}^{-1} \mathbf{y} \equiv \mathbf{q}_L^T \mathbf{y}$. □

In summary, Proposition 7 says that the "minimum" eigenvector of the *disturbance* autocorrelation matrix when used as the embedding signature allows the output SINR to attain its maximum possible value $\frac{\mathcal{D}_j}{\lambda_L + \sigma_n^2}$. At the same time, maximum SINR filtering becomes plain signature (eigenvector) matched filtering.

For fixed bit amplitude values $A_i = \sqrt{\mathcal{D}_i}$ and arbitrary signature initialization $\mathbf{s}_i \in \mathbb{R}^L$, $\|\mathbf{s}_i\| = 1$, $i = 1, 2, \ldots, K$, consider repeated applications of Proposition 7 for $j = 1, 2, \ldots, K$. In such an eigen-update signature cycle each signature is replaced by the minimum-eigenvalue eigenvector of the disturbance autocorrelation matrix as seen by the message corresponding to that signature. Once all signatures are updated, a second update cycle may begin. The whole procedure may continue for a predetermined number of cycles m or until convergence:

$$\mathbf{s}_j(m) \leftarrow \text{min eigenvector}\left(\mathbf{R}_{/j}(m)\right), \; j = 1, 2, \ldots, K, \quad m = 1, 2, \ldots . \tag{42}$$

It can be proven that convergence of (42) is guaranteed and, as shown in a code-division-multiple-access (CDMA) literature context [24], at each cycle the generalized total squared correlation TSC_g of the signature set $\mathbf{S} \stackrel{\triangle}{=} [\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_K]$, $TSC_g(\mathbf{S}) \stackrel{\triangle}{=} trace\left[\left(\mathbf{S}\mathbf{A}^2\mathbf{S}^T + \mathbf{R}_x + \sigma_n^2\mathbf{I}_L\right)^2\right]$ where $\mathbf{A} \stackrel{\triangle}{=} diag\,(A_1, A_2, \ldots A_K)$, is non-increasing [3].

If we consider channel coding the message bits before embedding and assume that the host $\mathbf{x}$ in (34) is Gaussian, the steganographic system determined by the signature matrix $\mathbf{S}$ is a special case of the K-user Gaussian multiple access channel with average input distortion constraints [23], [27]. For such a channel, the sum capacity C_{sum} (defined as the maximum sum of message coding rates at which messages can be recovered reliably [27], [28]) is a reasonable criterion of quality for the signature set $\mathbf{S}$ and equals

$$C_{sum}(\mathbf{S}, \mathbf{A}^2, \mathbf{R}_x, \sigma_n^2) = \frac{1}{2}\log\det\left(\mathbf{I}_L + \left(\mathbf{R}_x + \sigma_n^2\mathbf{I}_L\right)^{-1}\mathbf{S}\mathbf{A}^2\mathbf{S}^T\right). \tag{43}$$

Minimization of the $TSC_g(\mathbf{S})$ metric translates to maximization of C_{sum} [28]. However, decrease in TSC_g at each cycle of the algorithm in (42) does not necessarily imply increase of C_{sum} as seen for instance in [29],[30] via binary signature examples.

[3] Yet, there is no guarantee that $TSC_g(\mathbf{S})$ will converge to its minimum possible value (global minimum) [25], [26].

Apart from global optimality limitations, the multi-cycle multi-signature optimization procedure in (42) requires re-calculation of the disturbance autocorrelation matrix and eigen decomposition at each step of each cycle. A simple low-cost alternative to (42) could be a conditionally optimal single-cycle design method based on the following proposition.

Proposition 8 *Consider additive SS embedding according to*

$$\mathbf{y} = A_{l+1} b_{l+1} \mathbf{s}_{l+1} + \sum_{i=1}^{l} A_i b_i \mathbf{s}_i + \mathbf{x} + \mathbf{n}. \tag{44}$$

Let $\mathbf{q}_1, \mathbf{q}_2, \cdots \mathbf{q}_L$ *be eigenvectors of* $\mathbf{R}_x$ *in (1) with corresponding eigenvalues* $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_L$*. If the signatures* $\mathbf{s}_i$*,* $i = 1, 2, \ldots, l < L$*, are eigenvectors of* $\mathbf{R}_x$*, then for any watermark induced distortion level* $\mathcal{D}_{l+1} = A_{l+1}^2$ *a signature* $\mathbf{s}_{l+1}$ *that maximizes the output SINR of the maximum SINR filter for the bit of interest* b_{l+1} *subject to the constraint* $\mathbf{s}_{l+1}^T \mathbf{s}_i = 0$*,* $i = 1, 2, \ldots l$*, is*

$$\mathbf{s}_{l+1}^{opt} = \arg \max_{\substack{\mathbf{s}_{l+1} \\ \mathbf{s}_{l+1}^T \mathbf{s}_i = 0}} \left\{ A_{l+1}^2 \mathbf{s}_{l+1}^T \left(\mathbf{R}_x + \sigma_n^2 \mathbf{I}_L + \sum_{i=1}^{l} A_i b_i \mathbf{s}_i \right)^{-1} \mathbf{s}_{l+1} \right\} = \mathbf{q}_j \tag{45}$$

where $\mathbf{q}_j$ *is the minimum-eigenvalue eigenvector of* $\mathbf{R}_x$ *available.*

When $\mathbf{s}_{l+1} = \mathbf{q}_j$*, the output SINR is (conditionally) maximized to*

$$SINR_{max}(\mathbf{q}_j) = \frac{A_{l+1}^2}{\lambda_j + \sigma_n^2} = \frac{\mathcal{D}_{l+1}}{\lambda_j + \sigma_n^2} \tag{46}$$

and maximum SINR data filtering simplifies to

$$\mathbf{w}_{maxSINR}^T \mathbf{y} = \mathbf{q}_j^T (\mathbf{R}_x + \sigma_n^2 \mathbf{I}_L)^{-1} \mathbf{y} \equiv \mathbf{q}_j^T \mathbf{y}. \tag{47}$$

If, in addition, we are allowed to assume that $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_x)$*, then the optimum detector for the bit of interest* b_{l+1} *is* $\hat{b}_{l+1} = sign(\mathbf{q}_j^T \mathbf{y})$ *with probability of error* $P_e(\mathbf{q}_j) = Q\left(\sqrt{\frac{\mathcal{D}_{l+1}}{\lambda_j + \sigma_n^2}} \right)$*.* □

As a simple illustration of the use of Proposition 8 for conditionally optimal multi-signature design, suppose that we want to embed in the host data vector $\mathbf{x} \in \mathbb{R}^L$, $K \leq L$ message bits $b_1, b_2, \ldots, b_K$ with fixed corresponding amplitudes $A_1, A_2, \ldots, A_K$ (mean squared distortions $\mathcal{D}_1 = A_1^2$, $\mathcal{D}_2 = A_2^2, \ldots,$ $\mathcal{D}_K = A_K^2$) and signatures $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_K$ to be chosen. By Proposition 1 of Section II, we set $\mathbf{s}_1$ equal to the bottom eigenvector of $\mathbf{R}_x$, $\mathbf{s}_1 = \mathbf{q}_L$. Given $\mathbf{s}_1 = \mathbf{q}_L$ and under the constraint that we search for an $\mathbf{s}_2$ orthogonal to $\mathbf{s}_1$, by Proposition 8 we design $\mathbf{s}_2 = \mathbf{q}_{L-1}$ which is the next available eigenvector of $\mathbf{R}_x$ from the bottom. Given $\mathbf{s}_1 = \mathbf{q}_L$, $\mathbf{s}_2 = \mathbf{q}_{L-1}$ and under the constraint that we search for an $\mathbf{s}_3$ orthogonal to both $\mathbf{s}_2$ and $\mathbf{s}_1$, by Proposition 8 we assign $\mathbf{s}_3 = \mathbf{q}_{L-2}$. We continue this way until the

final assignment $\mathbf{s}_K = \mathbf{q}_{L-K+1}$. Once again, a welcome side effect of this conditionally SINR optimal signature design procedure is that the maximum SINR receiver for each message bit b_i, $i = 1, 2, \ldots, K$, simplifies to a matched filter and requires no knowledge of other system parameters.

For (fixed) unequal embedding amplitude values $A_1, A_2, \ldots, A_K$, the exact order by which the eigenvectors of $\mathbf{R}_x$ are drawn to become signatures is important if we consider the sum capacity of the steganographic system. We can show that a necessary condition for a maximum sum capacity solution under the constraint of eigenvector assignment is that the ordering of the bit amplitudes be inversely proportional to the ordering of the eigenvalues of the corresponding signature eigenvectors. This statement is given below in the form of a lemma. The proof can be found in the Appendix.

Lemma 1 *Consider additive SS embedding according to (34) with $K \leq L$ and let $\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_L$ be eigenvectors of $\mathbf{R}_x$ with corresponding eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_L$. Without loss of generality assume that $0 \leq A_1 \leq A_2 \leq \cdots \leq A_K$. Then,*

$$C_{sum}\left(\mathbf{s}_K = \mathbf{q}_L, \ldots, \mathbf{s}_{K-i} = \mathbf{q}_{L-i}, \ldots, \mathbf{s}_1 = \mathbf{q}_{L-K+1}\right) \geq$$
$$C_{sum}\left(\mathbf{s}_K = \mathbf{q}_L, \ldots, \mathbf{s}_{K-i} = \mathbf{q}_{L-j}, \ldots, \mathbf{s}_{K-j} = \mathbf{q}_{L-i}, \mathbf{s}_1 = \mathbf{q}_{L-K+1}\right), \; i, j \in \{0, 1, \ldots, K-1\}. \quad \square$$

If we generalize our approach and view the individual amplitudes/distortions as design parameters themselves, then we can search for the optimal amplitude assignment that maximizes sum capacity subject to a total allowable distortion constraint $\mathcal{D}_T = \sum_{i=1}^{L} A_i^2$. We derive the optimal amplitude values in the lemma below. The proof is given in the Appendix.

Lemma 2 *Consider additive SS embedding according to (34). Let $\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_L$ be eigenvectors of $\mathbf{R}_x$ with corresponding eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_L$. If the signatures $\mathbf{s}_i$ associated with the message bits b_i are distinct eigenvectors of $\mathbf{R}_x$, $\mathbf{s}_i = \mathbf{q}_{v_i}$, $v_i \in \{1, \cdots, L\}$, $i = 1, 2, \cdots, K$, then the sum capacity is maximized subject to an expected total distortion constraint $\mathcal{D}_T$ if*

$$A_i^2 = \left(-\left(\lambda_{v_i} + \sigma_n^2\right) + \mu\right)^+ \tag{48}$$

where $(x)^+ \stackrel{\triangle}{=} max(x, 0)$ and μ is the Kuhn-Tucker coefficient [23] chosen such that the distortion constraint $\mathcal{D}_T = \sum_{i=1}^{K} A_i^2$ is met. $\square$

To find the necessary parameter value μ in (48) we suggest to arrange the participating eigenvalues λ_{v_i}, $i = 1, 2, \cdots, K$, in ascending order: $\lambda_{z_1} \leq \lambda_{z_2} \leq \cdots \leq \lambda_{z_K}$. Then,

$$\mu = \frac{\sum_{i=1}^{J^*}(\lambda_{z_i} + \sigma_n^2) + \mathcal{D}_T}{J^*} \tag{49}$$

where the cutoff index J^* is the greatest integer in $\left\{J : J \in \{1,2,\cdots,K\} \text{ and } \frac{\sum_{i=1}^{J}(\lambda_{z_i}+\sigma_n^2)+\mathcal{D}_T}{J} > \lambda_{z_J}+\sigma_n^2\right\}$. The optimal message amplitude/distortion allocation solution of Lemma 2 can be viewed as a power waterfilling procedure [23] in the eigen domain of the host.

Finally, as the last technical development in this paper we examine the possibility of carrying out multi-signature embedding in linearly transformed host data. We assume that the host data vector $\mathbf{x}$ is linearly transformed by an $L \times L$ operator of the form $\mathbf{I}_L - \sum_{i=1}^{K} c_i \mathbf{s}_i \mathbf{s}_i^T$ where $c_i \in \mathbb{R}$ and $\mathbf{s}_i \in \mathbb{R}^L$, $\|\mathbf{s}_i\| = 1$, $i = 1,2,\cdots,K$, are the parameters and signatures to be designed. The final composite signal is

$$\mathbf{y} = \sum_{i=1}^{K} A_i b_i \mathbf{s}_i + \left(\mathbf{I}_L - \sum_{i=1}^{K} c_i \mathbf{s}_i \mathbf{s}_i^T\right)\mathbf{x} + \mathbf{n} \tag{50}$$

and the mean squared distortion induced by each individual message i, $i = 1,2,\ldots,K$, is

$$\mathcal{D}_i = E\left\{\left\|\left(A_i b_i - c_i \mathbf{s}_i^T \mathbf{x}\right)\mathbf{s}_i\right\|^2\right\} = A_i^2 + c_i^2 \mathbf{s}_i^T \mathbf{R}_x \mathbf{s}_i. \tag{51}$$

With signal of interest $A_j b_j \mathbf{s}_j$, the autocorrelation matrix of the disturbance is $\mathbf{R}_{/j} = \sum_{\substack{i=1 \\ i\neq j}}^{K} {A_i}^2 \mathbf{s}_i {\mathbf{s}_i}^T + \left(\mathbf{I}_L - \sum_{i=1}^{K} c_i \mathbf{s}_i \mathbf{s}_i^T\right)\mathbf{R}_x\left(\mathbf{I}_L - \sum_{i=1}^{K} c_i \mathbf{s}_i \mathbf{s}_i^T\right) + \sigma_n^2 \mathbf{I}_L$. Unfortunately, in contrast to (38) for multi-signature embedding in non-transformed data, $\mathbf{R}_{/j}$ remains a function of $\mathbf{s}_j$ (as well as c_j). In this context, unconditionally optimal multi-signature multi-cycle optimization along the lines of (42) is practically an unrealistic objective. Instead, we suggest to design sequentially the amplitudes A_i, parameters c_i, and signatures $\mathbf{s}_i$ of the embedded messages $i = 1,2,\ldots,K \leq L$, such that the output SINR is conditionally maximized given all past fixed embeddings (single-cycle optimization). Our developments are presented in the form of Proposition 9 below whose proof is given in the Appendix.

Proposition 9 *Consider additive SS embedding according to*

$$\mathbf{y} = A_{l+1} b_{l+1} \mathbf{s}_{l+1} + \sum_{i=1}^{l} A_i b_i \mathbf{s}_i + \left(\mathbf{I}_L - \sum_{i=1}^{l} c_i \mathbf{s}_i \mathbf{s}_i^T - c_{l+1}\mathbf{s}_{l+1}\mathbf{s}_{l+1}^T\right)\mathbf{x} + \mathbf{n}. \tag{52}$$

Let $\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_L$ be eigenvectors of $\mathbf{R}_x$ in (1) with corresponding eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_L$. Assume that $\mathbf{s}_i$, $i = 1,2,\ldots,l < L$, are all eigenvectors of $\mathbf{R}_x$ and j is the index of the minimum-eigenvalue eigenvector of $\mathbf{R}_x$ available. For any given message induced distortion level $\mathcal{D}_{l+1}$, an $(\mathbf{s}_{l+1}, c_{l+1})$ pair that maximizes the output SINR of the maximum SINR filter subject to the constraint $\mathbf{s}_{l+1}^T \mathbf{s}_i = 0$, $i = 1,2,\ldots,l$, is

$$\begin{aligned}\mathbf{s}_{l+1}^{opt} \;=\; \arg\max_{\substack{\mathbf{s}_{l+1} \\ \mathbf{s}_{l+1}^T \mathbf{s}_i = 0}} \Bigg\{ & A_{l+1}^2 \mathbf{s}_{l+1}^T \Bigg(\sum_{i=1}^{l} {A_i}^2 \mathbf{s}_i {\mathbf{s}_i}^T + \\ & \left(\mathbf{I}_L - \sum_{i=1}^{l} c_i \mathbf{s}_i \mathbf{s}_i^T - c_{l+1}\mathbf{s}_{l+1}\mathbf{s}_{l+1}^T\right)\mathbf{R}_x\left(\mathbf{I}_L - \sum_{i=1}^{l} c_i \mathbf{s}_i \mathbf{s}_i^T - c_{l+1}\mathbf{s}_{l+1}\mathbf{s}_{l+1}^T\right) + \sigma_n^2 \mathbf{I}_L\Bigg)^{-1} \Bigg\} = \mathbf{q}_j \end{aligned} \tag{53}$$

and $c_{l+1}^{opt} = \arg\max_{c_{l+1}} \left\{ SINR_{max}(\mathbf{s}_{l+1}^{opt}, c_{l+1}) \right\} = \dfrac{\lambda_j + \sigma_n^2 + \mathcal{D}_{l+1} - \sqrt{\left(\lambda_j + \sigma_n^2 + \mathcal{D}_{l+1}\right)^2 - 4\lambda_j \mathcal{D}_{l+1}}}{2\lambda_j}$. (54)

When $\mathbf{s}_{l+1} = \mathbf{q}_j$ *and* $c_{l+1} = c_{l+1}^{opt}$, *the output SINR is (conditionally) maximized to*

$$SINR_{max}(\mathbf{q}_j, c_{l+1}^{opt}) = \frac{\mathcal{D}_{l+1} - c_{l+1}^{opt\,2}\lambda_j}{\lambda_j \left(1 - c_{l+1}^{opt}\right)^2 + \sigma_n^2} \tag{55}$$

and maximum SINR data filtering simplifies to $\mathbf{q}_j^T \mathbf{y}$. *If, in addition, we are allowed to assume that* $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_x)$, *then the optimum detector for the bit of interest* b_{l+1} *is* $\hat{b}_{l+1} = sgn(\mathbf{q}_j^T \mathbf{y})$ *with probability of error*

$$P_e(\mathbf{q}_j, c_{l+1}^{opt}) = Q\left(\sqrt{\frac{\mathcal{D}_{l+1} - c_{l+1}^{opt\,2}\lambda_j}{\lambda_j \left(1 - c_{l+1}^{opt}\right)^2 + \sigma_n^2}} \right). \quad \square \tag{56}$$

As an illustrative example of the use of Proposition 9, suppose that we would like to embed in the host data vector $\mathbf{x}$, $K \leq L$ message bits $b_1, b_2, \ldots, b_K$ with individual corresponding mean squared host distortion $\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_K$. We first design the system parameters $\mathbf{s}_1$, c_1, and A_1 for message bit b_1 in the absence of any other message in the host image. According to Proposition 9 ($l = 0$), the optimal parameter selection is $\mathbf{s}_1 = \mathbf{q}_L$, $c_1 = \frac{\lambda_L + \sigma_n^2 + \mathcal{D}_1 - \sqrt{\left(\lambda_L + \sigma_n^2 + \mathcal{D}_1\right)^2 - 4\lambda_L \mathcal{D}_1}}{2\lambda_L}$, and by (51) $A_1 = \sqrt{\mathcal{D}_1 - c_1^2 \lambda_L}$. Next, we proceed with the second message bit b_2 and optimize $\mathbf{s}_2$ and c_2 subject to the desired distortion level $\mathcal{D}_2$ and the constraint $\mathbf{s}_2^T \mathbf{s}_1 = 0$. Since $\mathbf{s}_1$ is already an eigenvector of $\mathbf{R}_x$, Proposition 9 ($l = 1$) offers $\mathbf{s}_2 = \mathbf{q}_{L-1}$, $c_2 = \frac{\lambda_{L-1} + \sigma_n^2 + \mathcal{D}_2 - \sqrt{\left(\lambda_{L-1} + \sigma_n^2 + \mathcal{D}_2\right)^2 - 4\lambda_{L-1}\mathcal{D}_2}}{2\lambda_{L-1}}$, and by (51) $A_2 = \sqrt{\mathcal{D}_2 - c_2^2 \lambda_{L-1}}$. We continue calculating signatures $\mathbf{s}_i$, parameters c_i, and amplitudes A_i as above, always subject to the desired distortion $\mathcal{D}_i$ and orthogonality between the signature to be designed and all other prior signatures. Provided that $K \leq L$, the final set of designed parameters is $\mathbf{s}_K = \mathbf{q}_{L-(K-1)}$, $c_K = \frac{\lambda_{L-(K-1)} + \sigma_n^2 + \mathcal{D}_K - \sqrt{\left(\lambda_{L-(K-1)} + \sigma_n^2 + \mathcal{D}_K\right)^2 - 4\lambda_{L-(K-1)}\mathcal{D}_K}}{2\lambda_{L-(K-1)}}$, and $A_K = \sqrt{\mathcal{D}_K - c_K^2 \lambda_{L-(K-1)}}$.

Optimal distortion allocation for sum capacity maximization in multi-message embedding in linearly transformed host data subject to the constraint that all messages are assigned distinct eigenvectors of $\mathbf{R}_x$, $\mathbf{s}_i = \mathbf{q}_{v_i}$, and subject to a total distortion constraint $\mathcal{D}_T = \sum_{i=1}^K \mathcal{D}_i = \sum_{i=1}^K \left(A_i^2 + c_i \lambda_{v_i}\right)$, $v_i \in \{1, \cdots, L\}$, $i = 1, 2, \ldots K$, is a joint optimization problem with respect to c_i and A_i. We suggest an iterative solution approach based on Proposition 9 above and Lemma 2 presented earlier in this section. We initially fix the distortions induced by each message and find the optimum c_i parameters according to Proposition 9. Then, we perform optimum amplitude allocation according to Lemma 2. Based on this allocation, we re-evaluate all c_i by Proposition 9. We continue until convergence is observed.

(a) (b)

Fig. 2. (a) Aircraft image example (512×512 grey-scale). (b) Aircraft image after $20dB$ ($\mathbf{s}^{\text{opt}}$, c^{opt}) embedding of $4Kbits$ and additive white Gaussian noise of variance $3dB$.

4. Experimental Studies

To carry out an experimental study of the technical developments presented in the previous sections, we consider as a host example the familiar grey scale 512×512 "Aircraft" image in Fig. 2(a) that has been used widely in the pertinent literature. We perform 8×8 block DCT single-signature embedding over all bins except the dc coefficient. Hence, our signature length is $L = 63$ and we embed $\frac{512^2}{8^2} = 4,096$ bits. For the sake of generality, we also incorporate white Gaussian noise of variance $\sigma_n^2 = 3$dB. Fig. 3 shows the recovery bit-error-rate (BER) under signature matched filter detection as a function of the distortion created by the embedded message over the 0 to $20dB$ range for four different embedding schemes: *(a)* SS embedding with an arbitrary signature, *(b)* SS embedding with an arbitrary signature and optimized selection of the host data transformation parameter c as in [18] (known as "improved spread-spectrum" or ISS), *(c)* SS embedding with an optimal signature according to Proposition 1, and *(d)* SS embedding with a jointly optimal signature and host data transformation parameter ($\mathbf{s}^{\text{opt}}$, c^{opt}) according to Proposition 4. The demonstrated BER improvement of our joint signature and parameter optimization procedure in particular, measures in orders of magnitude. Fig. 2(b) shows the Aircraft image after $20dB$ ($\mathbf{s}^{\text{opt}}$, c^{opt})-embedding of the $4Kbit$ message and $3dB$ additive white Gaussian noise disturbance in the block DCT domain.

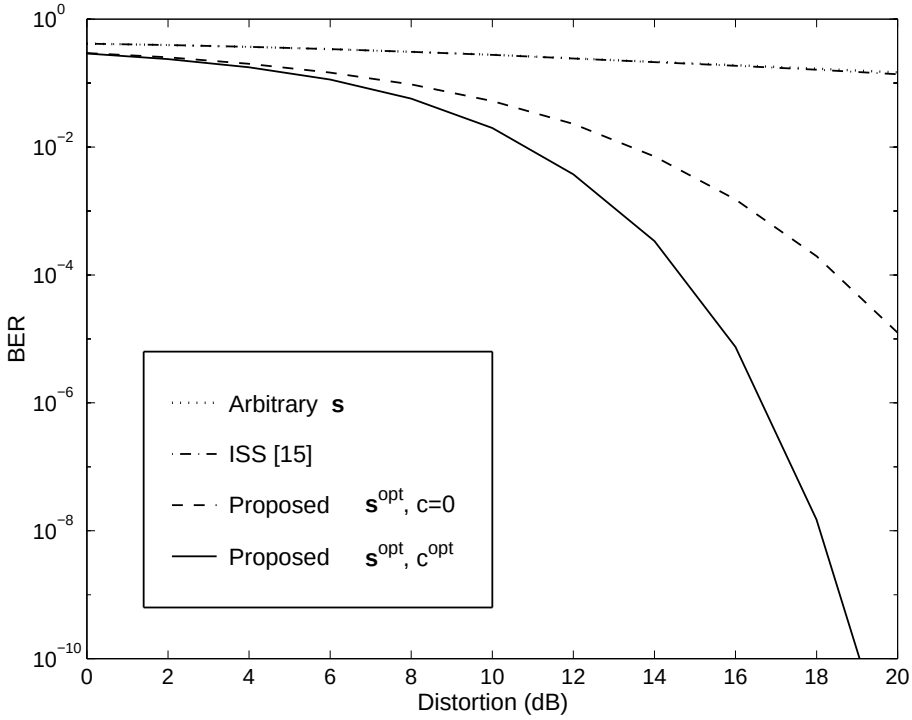


Fig. 3. Bit-error-rate versus host distortion (Aircraft image, $\sigma_n^2 = 3dB$).

In Fig. 4, we repeat the same experiment of Fig. 3 on the 256×256 grey scale Baboon image in Fig. 1(a) (signature length $L = 63$, hidden message of $\frac{256^2}{8^2} = 1,024$ bits, and additive white Gaussian noise disturbance of variance $3dB$). Comparatively speaking (Aircraft versus Baboon host or Fig. 3 versus Fig. 4 results), message recovery for the Baboon host appears to be a somewhat more difficult problem. Yet, the proposed joint signature and host transformation parameter optimization scheme maintains a near 10^{-10} BER at $20dB$ host distortion and outperforms the proposed signature-only optimization scheme by about eight orders of magnitude.

In Fig. 5, we continue our experimental work with the Baboon host and plot the capacity versus distortion performance curves for the four embedders under consideration. We see, for example, that at $20dB$ host distortion the jointly optimized embedder offers 0.8 information bit payload per embedded bit. This number goes down to 0.35 for signature only optimization, 0.18 for ISS embedding [18], and 0.15 for arbitrary signature embedding.

Next, we consider the problem of multi-signature embedding. We keep the Baboon image as the host and wish to hide $K = 15$ data blocks/messages of length $1,024$ bits each with each block/message having its own individual embedding signature. Each message is allowed to cause the same expected distortion to the host $\mathcal{D}_i = D$, $i = 1, \ldots, K$. Therefore, for statistically independent messages, the total distortion

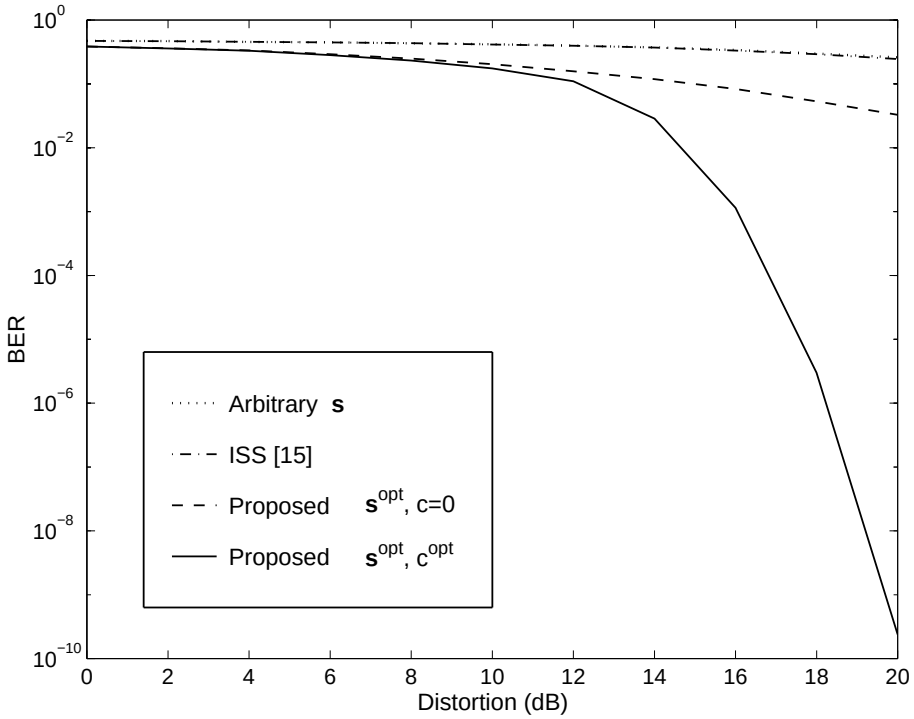


Fig. 4. Bit-error-rate versus host distortion (Baboon image, $\sigma_n^2 = 3dB$).

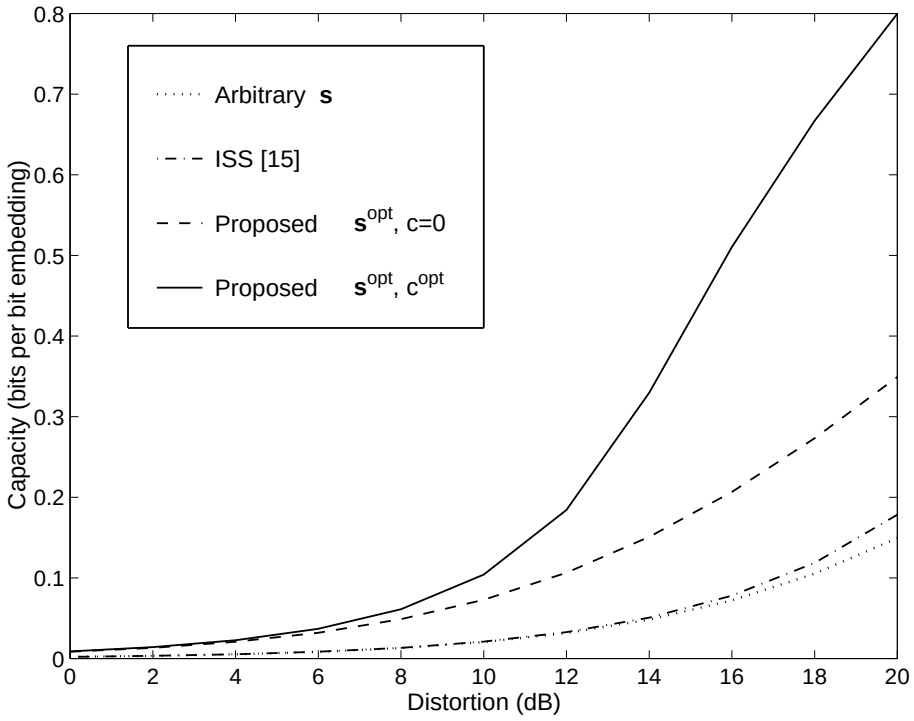


Fig. 5. Capacity versus distortion (Baboon image, $\sigma_n^2 = 3dB$).

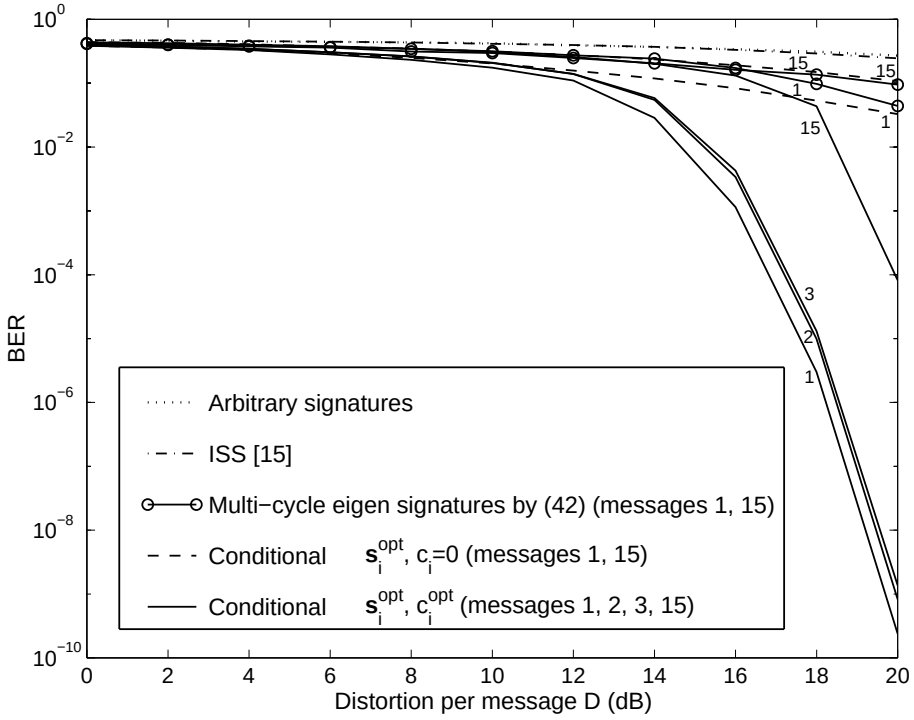


Fig. 6. BER as a function of the per-message distortion D (baboon image, $K = 15$ messages of size $1,024$ bits each, $\sigma_n^2 = 3dB$).

to the host is $\sum_{i=1}^{K} \mathcal{D}_i = KD$. As before, for the sake of generality, we add to the host white Gaussian noise of variance $3dB$. We study five different multi-signature embedding schemes: *(a)* Embedding with arbitrary signatures, *(b)* ISS embedding [18], *(c)* multi-cycle eigen-signature design by (42), *(d)* conditional optimization by Proposition 8 (sequential $\mathbf{s}_i^{\rm opt}$, $c_i = 0$ assignment, $i = 1, \ldots, 15$), and *(e)* conditional optimization by Proposition 9 (sequential $\mathbf{s}_i^{\rm opt}$, $c_i^{\rm opt}$ assignment, $i = 1, \ldots, 15$). As seen in Fig. 6, the superiority of the latter approach (design by Proposition 9) is evident. In fact, under joint sequential $\mathbf{s}_i^{\rm opt}$, $c_i^{\rm opt}$ optimization even the least favored message ($i = 15$) outperforms in recovery BER the most favored ($i = 1$) message under sequential $\mathbf{s}_i^{\rm opt}$, $c_i = 0$ signature-only optimization or multi-cycle eigen-signature design for per-message distortion values above $18dB$. Fig. 7 shows the Baboon image after embedding the fifteen messages ($15 \cdot 1,024$ bits) via joint sequential $\mathbf{s}_i^{\rm opt}$, $c_i^{\rm opt}$, $i = 1, \ldots, 15$, optimization with $20dB$ per-message distortion ($31.8dB$ total distortion) and $3dB$ variance additive white Gaussian noise.

Finally, in Fig. 8 we present sum capacity results when the two proposed schemes, sequential $\mathbf{s}_i^{\rm opt}$ design and sequential joint $\mathbf{s}_i^{\rm opt}$, $c_i^{\rm opt}$ design, $i = 1, \ldots, 15$, employ waterfilling power allocation (use of Lemma 2 alone or coupled use of Proposition 9 and Lemma 2, correspondingly). We see, for example,

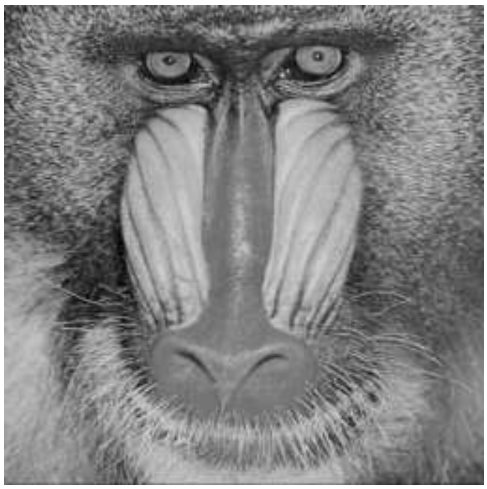


Fig. 7. Baboon image after multi-signature embedding via Proposition 9 ($K = 15$ messages of size $1,024$ bits each, per-message distortion $20dB$, $\sigma_n^2 = 3dB$).

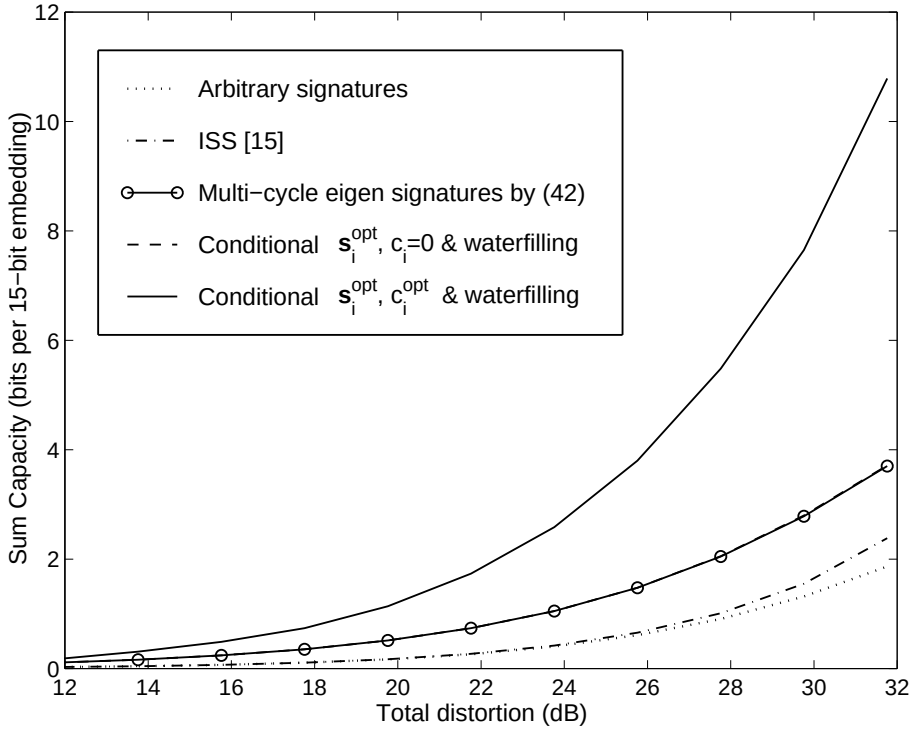


Fig. 8. Sum capacity versus distortion (baboon image, $K = 15$, $\sigma_n^2 = 3dB$).

that at $32dB$ total distortion the waterfilled $\mathbf{s}_i^{\text{opt}}$, c_i^{opt} design offers information bit payload of about 11 bits per 15 bits embedded, while the waterfilled $\mathbf{s}_i^{\text{opt}}$, $c_i = 0$ design offers only about 4 bits per 15 bits embedded.

5. Conclusion

We considered the problem of hiding digital data in a digital host image (or audio) via spread-spectrum embedding in an arbitrary transform domain. We showed that use of the minimum-eigenvalue eigenvector of the transform domain host data autocorrelation matrix as the embedding signature offers the maximum possible SINR under linear filter message recovery and, conveniently, does so under plain signature correlation (signature matched filtering). If we allow ourselves the added assumption of (colored) Gaussian transform-domain host data, then we see that the above described system as a whole becomes minimum probability of error and maximum Shannon capacity optimal as well.

To take these findings one step further, we examined SS embedding in transform-domain host data that are modified by a parametrized projection-like linear operator. We found the joint signature and parameter values under the optimality scenaria mentioned above. Conveniently, the jointly optimal signature is still the minimum-eigenvalue eigenvector and the SINR optimal linear filter at the receiver side is still the signature correlator. Yet, joint signature and parameter optimization was seen to offer dramatic improvements in SINR, probabilty of error, and capacity (Figs. 3, 4, and 5, for example).

Finally, we extended our effort to cover multi-signature/multi-message embedding. First, under signature-only optimization we developed a computationally costly multi-cycle eigen-signature design scheme based on the disturbance autocorrelation matrices. The alternative suggestion based on the host data autocorrelation matrix alone and sequential (conditional) eigen-signature optimization is practically much more appealing. A waterfilling amplitude assignment algorithm was developed as well to maximize sum capacity under eigen-signature designs. All multi-signature findings were generalized to cover parametrized projection-like modification of the host data with, once again, dramatic improvements in probabilty of error or sum capacity (as seen in Figs. 6 and 8, for example).

As a brief concluding remark, image-adaptive signature(s) or signature(s)/parameter(s) optimization as described in this work can be carried out over a set of host images (frames) if desired. The only technical difference is the calculation of the host data autocorrelation matrix which now has to extend over the whole host set. As long as the cumulative host autocorrelation matrix is not constant-value diagonal ($\neq \alpha\mathbf{I}$), significant gains are to be collected over standard non-adaptive SS embedding

techniques.

Appendix

Proof of Proposition 3

We need to find $\mathbf{s}$ that maximizes C in (15) subject to $\|\mathbf{s}\| = 1$. Since $\log(\cdot)$ is a strictly monotonic function,

$$\mathbf{s}^{\text{opt}} = \underset{\mathbf{s}\in\mathbb{R}^L,\ \|\mathbf{s}\|=1}{\arg\max}\ \det\left(\mathbf{I}_L + \mathcal{D}\left(\mathbf{R}_x + \sigma_n^2\mathbf{I}_L\right)^{-1}\mathbf{s}\mathbf{s}^T\right) = \underset{\mathbf{s}\in\mathbb{R}^L,\ \|\mathbf{s}\|=1}{\arg\max}\ \det\left(\frac{1}{\mathcal{D}}\left(\mathbf{R}_x + \sigma_n^2\mathbf{I}_L\right) + \mathbf{s}\mathbf{s}^T\right). \quad (57)$$

Using the rank-one update rule [33],
$\det\left(\frac{1}{\mathcal{D}}\left(\mathbf{R}_x + \sigma_n^2\mathbf{I}_L\right) + \mathbf{s}\mathbf{s}^T\right) = \frac{1}{\mathcal{D}^L}\det\left(\mathbf{R}_x + \sigma_n^2\mathbf{I}_L\right)\left(1 + \mathcal{D}\mathbf{s}^T\left(\mathbf{R}_x + \sigma_n^2\mathbf{I}_L\right)^{-1}\mathbf{s}\right)$. Therefore,

$$\mathbf{s}^{\text{opt}} = \underset{\mathbf{s}\in\mathbb{R}^L,\ \|\mathbf{s}\|=1}{\arg\max}\left(1 + \mathcal{D}\mathbf{s}^T\left(\mathbf{R}_x + \sigma_n^2\mathbf{I}_L\right)^{-1}\mathbf{s}\right) = \mathbf{q}_L. \quad \square \quad (58)$$

Proof of Proposition 4

For a target distortion value $\mathcal{D}$, the term A^2 in (21) equals $\mathcal{D} - c^2\mathbf{s}^T\mathbf{R}_x\mathbf{s}$ and is maximized for $\mathbf{s} = \mathbf{q}_L$. We will show that the second term in (21), $\mathbf{s}^T\left(\left(\mathbf{I}_L - c\mathbf{s}\mathbf{s}^T\right)\mathbf{R}_x\left(\mathbf{I}_L - c\mathbf{s}\mathbf{s}^T\right) + \sigma_n^2\mathbf{I}_L\right)^{-1}\mathbf{s}$, is maximized by $\mathbf{s} = \mathbf{q}_L$ as well. By the matrix inversion lemma,

$$\begin{aligned}
&\left(\left(\mathbf{I}_L - c\mathbf{s}\mathbf{s}^T\right)\mathbf{R}_x\left(\mathbf{I}_L - c\mathbf{s}\mathbf{s}^T\right) + \sigma_n^2\mathbf{I}_L\right)^{-1} \\
&= \frac{1}{\sigma_n^2}\mathbf{I}_L - \frac{1}{\sigma_n^4}\mathbf{I}_L\left(\mathbf{I}_L - c\mathbf{s}\mathbf{s}^T\right)\left(\mathbf{R}_x^{-1} + \frac{1}{\sigma_n^2}\left(\mathbf{I}_L - c\mathbf{s}\mathbf{s}^T\right)^2\right)^{-1}\left(\mathbf{I}_L - c\mathbf{s}\mathbf{s}^T\right) \text{ and} \quad (59)
\end{aligned}$$

$$\begin{aligned}
&\left(\mathbf{R}_x^{-1} + \frac{1}{\sigma_n^2}\left(\mathbf{I}_L - c\mathbf{s}\mathbf{s}^T\right)^2\right)^{-1} \\
&= \left(\mathbf{R}_x^{-1} + \frac{1}{\sigma_n^2}\mathbf{I}_L\right)^{-1} - \frac{\left(c^2 - 2c\right)\left(\mathbf{R}_x^{-1} + \frac{1}{\sigma_n^2}\mathbf{I}_L\right)^{-1}\mathbf{s}\mathbf{s}^T\left(\mathbf{R}_x^{-1} + \frac{1}{\sigma_n^2}\mathbf{I}_L\right)^{-1}}{\sigma_n^2 + \left(c^2 - 2c\right)\mathbf{s}^T\left(\mathbf{R}_x^{-1} + \frac{1}{\sigma_n^2}\mathbf{I}_L\right)^{-1}\mathbf{s}}. \quad (60)
\end{aligned}$$

Combining (59) and (60) we obtain

$$\mathbf{s}^T\left(\left(\mathbf{I}_L - c\mathbf{s}\mathbf{s}^T\right)\mathbf{R}_x\left(\mathbf{I}_L - c\mathbf{s}\mathbf{s}^T\right) + \sigma_n^2\mathbf{I}_L\right)^{-1}\mathbf{s} = \frac{1}{\sigma_n^2}\,\frac{\sigma_n^2 - \mathbf{s}^T\left(\mathbf{R}_x^{-1} + \frac{1}{\sigma_n^2}\mathbf{I}_L\right)^{-1}\mathbf{s}}{\sigma_n^2 + \left(c^2 - 2c\right)\mathbf{s}^T\left(\mathbf{R}_x^{-1} + \frac{1}{\sigma_n^2}\mathbf{I}_L\right)^{-1}\mathbf{s}}. \quad (61)$$

The derivative of the righthandside of (61) with respect to $\mathbf{s}^T\left(\mathbf{R}_x^{-1} + \frac{1}{\sigma_n^2}\mathbf{I}_L\right)^{-1}\mathbf{s}$ gives

$$\frac{1}{\sigma_n^2}\frac{-\sigma_n^2\left(1 + c\left(c - 2\right)\right)}{\left(\sigma_n^2 + c\left(c - 2\right)\mathbf{s}^T\left(\mathbf{R}_x^{-1} + \frac{1}{\sigma_n^2}\mathbf{I}_L\right)^{-1}\mathbf{s}\right)^2} \le 0 \quad \forall c \in \mathbb{R}. \quad (62)$$

Hence, $\mathbf{s}^T\left(\left(\mathbf{I}_L - c\mathbf{s}\mathbf{s}^T\right)\mathbf{R}_x\left(\mathbf{I}_L - c\mathbf{s}\mathbf{s}^T\right) + \sigma_n^2\mathbf{I}_L\right)^{-1}\mathbf{s}$ is a decreasing function of $\mathbf{s}^T\left(\mathbf{R}_x^{-1} + \frac{1}{\sigma_n^2}\mathbf{I}_L\right)^{-1}\mathbf{s}$. Yet, $\frac{\lambda_L}{1+\frac{\lambda_L}{\sigma_n^2}} \le \mathbf{s}^T\left(\mathbf{R}_x^{-1} + \frac{1}{\sigma_n^2}\mathbf{I}_L\right)^{-1}\mathbf{s} \le \frac{\lambda_1}{1+\frac{\lambda_1}{\sigma_n^2}}$. Therefore,

$$\arg\max_{\mathbf{s}}\left\{\mathbf{s}^T\left(\left(\mathbf{I}_L - c\mathbf{s}\mathbf{s}^T\right)\mathbf{R}_x\left(\mathbf{I}_L - c\mathbf{s}\mathbf{s}^T\right) + \sigma_n^2\mathbf{I}_p\right)^{-1}\mathbf{s}\right\} = \mathbf{q}_L. \quad (63)$$

Since $\mathbf{s}^{\text{opt}} = \mathbf{q}_L$ for any c,

$$\max_{c,\mathbf{s}} \{\text{SINR}_{\max}\} = \max_{c} \left\{ \max_{\mathbf{s}} \{\text{SINR}_{\max}\} \right\} = \max_{c} \left\{ \frac{\mathcal{D} - c^2\lambda_L}{\sigma_n^2 + \lambda_L + c(c-2)\lambda_L} \right\}. \tag{64}$$

By direct differentiation of the final expression in (64) and root selection, we obtain c^{opt} in (23). □

Proof of Proposition 6

For the signal model in (18), the channel capacity is

$$C = \max_{f_b : E\{b^2\} \leq \frac{\mathcal{D} - c^2 \mathbf{s}^T \mathbf{R}_x \mathbf{s}}{A^2}} I(b;\mathbf{y}) = \frac{1}{2} \log \frac{\det\left(\left(\mathcal{D} - c^2 \mathbf{s}^T \mathbf{R}_x \mathbf{s}\right) \mathbf{s}\mathbf{s}^T + \left(\mathbf{I}_L - c\mathbf{s}\mathbf{s}^T\right) \mathbf{R}_x \left(\mathbf{I}_L - c\mathbf{s}\mathbf{s}^T\right) + \sigma_n^2 \mathbf{I}_L \right)}{\det\left(\left(\mathbf{I}_L - c\mathbf{s}\mathbf{s}^T\right) \mathbf{R}_x \left(\mathbf{I}_L - c\mathbf{s}\mathbf{s}^T\right) + \sigma_n^2 \mathbf{I}_L \right)}. \tag{65}$$

Applying a rank-one update [33] to the determinant in the numerator, we obtain

$$C = \frac{1}{2} \log \left(1 + \left(\mathcal{D} - c^2 \mathbf{s}^T \mathbf{R}_x \mathbf{s}\right) \mathbf{s}^T \left(\left(\mathbf{I}_L - c\mathbf{s}\mathbf{s}^T\right) \mathbf{R}_x \left(\mathbf{I}_L - c\mathbf{s}\mathbf{s}\right) + \sigma_n^2 \mathbf{I}_L \right)^{-1} \mathbf{s} \right). \tag{66}$$

The result follows from the proof of Proposition 4. □

Proof of Lemma 1

The sum capacity of the channel in (34) provided that all signatures $\mathbf{s}_i$ associated with the message bits b_i are distinct eigenvectors of $\mathbf{R}_x$ ($\mathbf{s}_i = \mathbf{q}_{v_i}$, $v_i \in \{1, 2, \cdots, L\}$, $i = 1, 2, \cdots, K$) is given by

$$C_{sum} = \frac{1}{2} \log \prod_{j=1}^{K} \left(1 + \frac{A_j^2}{\lambda_{v_j} + \sigma_n^2} \right). \tag{67}$$

Then, $C_{sum}\left(\mathbf{s}_K = \mathbf{q}_L, \ldots, \mathbf{s}_{K-i} = \mathbf{q}_{L-i}, \ldots, \mathbf{s}_1 = \mathbf{q}_{L-K+1}\right) - C_{sum}\left(\mathbf{s}_K = \mathbf{q}_L, \ldots, \mathbf{s}_{K-i} = \mathbf{q}_{L-j}, \ldots, \mathbf{s}_{K-j} = \mathbf{q}_{L-i}, \ldots, \mathbf{s}_1 = \mathbf{q}_{L-K+1}\right) = \frac{1}{2} \log \frac{\left(1 + \frac{A_{K-i}^2}{\lambda_{L-i} + \sigma_n^2}\right)\left(1 + \frac{A_{K-j}^2}{\lambda_{L-j} + \sigma_n^2}\right)}{\left(1 + \frac{A_{K-i}^2}{\lambda_{L-j} + \sigma_n^2}\right)\left(1 + \frac{A_{K-j}^2}{\lambda_{L-i} + \sigma_n^2}\right)}$. But, $\left(1 + \frac{A_{K-i}^2}{\lambda_{L-i} + \sigma_n^2}\right)\left(1 + \frac{A_{K-j}^2}{\lambda_{L-j} + \sigma_n^2}\right) - \left(1 + \frac{A_{K-i}^2}{\lambda_{L-j} + \sigma_n^2}\right)\left(1 + \frac{A_{K-j}^2}{\lambda_{L-i} + \sigma_n^2}\right) = \frac{A_{K-i}^2 - A_{K-j}^2}{\lambda_{L-i} + \sigma_n^2} - \frac{A_{K-i}^2 - A_{K-j}^2}{\lambda_{L-j} + \sigma_n^2} \leq 0$ for $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_L$ and $A_1 \leq A_2 \leq \cdots \leq A_K$. □

Proof of Lemma 2

To identify the set of amplitudes (or equivalently distortions) that maximizes the concave sum capacity function in (67) subject to the distortion constraint $\mathcal{D}_T$, we differentiate the Lagrange functional

$$\mathcal{L}(A_1, A_2, \cdots, A_K) = \frac{1}{2} \sum_{j=1}^{K} \log \left(1 + \frac{A_j^2}{\lambda_{v_j} + \sigma_n^2} \right) - \mu_1 \sum_{j=1}^{K} A_j^2 \tag{68}$$

with respect to A_j and obtain

$$A_j^2 = \left(- \left(\lambda_{v_j} + \sigma_n^2\right) + \mu \right)^+ \tag{69}$$

where $\mu = \frac{1}{2\mu_1}$ is the Kuhn-Tucker coefficient [23] chosen such that $\mathcal{D}_T = \sum_{i=1}^{K} A_i^2$. Substitution of the optimal amplitude allocation of (69) in (67) gives the maximum attainable sum capacity value

$C_{sum} = \frac{1}{2}\log\left[\mu^{J^*}/\prod_{j=1}^{J^*}(\lambda_{z_j}+\sigma_n^2)\right]$ where μ, J^*, and λ_{z_j} are as in (49). □

Proof of Proposition 9

The output SINR of the maximum SINR filter for the signal model in (52) is

$$\begin{aligned}&\mathrm{SINR}_{\max} = \left(\mathcal{D}_{l+1} - c_{l+1}^2 \mathbf{s}_{l+1}^T \mathbf{R}_x \mathbf{s}_{l+1}\right) \cdot \\ &\mathbf{s}_{l+1}^T \left(\left(\mathbf{I}_L - \sum_{i=1}^{l} c_i \mathbf{s}_i \mathbf{s}_i^T - c_{l+1}\mathbf{s}_{l+1}\mathbf{s}_{l+1}^T\right)\mathbf{R}_x \left(\mathbf{I}_L - \sum_{i=1}^{l} c_i \mathbf{s}_i \mathbf{s}_i^T - c_{l+1}\mathbf{s}_{l+1}\mathbf{s}_{l+1}^T\right) + \sigma_n^2 \mathbf{I}_L + \sum_{i=1}^{l} A_i \mathbf{s}_i \mathbf{s}_i^T\right)^{-1} \mathbf{s}_{l+1}.\end{aligned} \tag{70}$$

Let $\mathbf{Q} \stackrel{\triangle}{=} [\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_L]$ be the matrix with columns the eigenvectors of $\mathbf{R}_x$, $\mathbf{\Lambda} \stackrel{\triangle}{=} diag\,(\lambda_1, \lambda_2, \ldots, \lambda_L)$ the diagonal matrix with the eigenvalues of $\mathbf{R}_x$, and p the number of available eigenvectors of $\mathbf{R}_x$ that do not correspond to any $\mathbf{s}_i$, $i = 1, \ldots, l$. Since $\mathbf{s}_i$, $i = 1, \ldots, l$, are eigenvectors of $\mathbf{R}_x$ we can write

$$\sigma_n^2 \mathbf{I}_L + \sum_{i=1}^{l} A_i \mathbf{s}_i \mathbf{s}_i^T = \mathbf{Q}\mathbf{\Lambda}_1\mathbf{Q}^T \text{ and } \left(\mathbf{I}_L - \sum_{i=1}^{l} c_i \mathbf{s}_i \mathbf{s}_i^T\right)\mathbf{R}_x\left(\mathbf{I}_L - \sum_{i=1}^{l} c_i \mathbf{s}_i \mathbf{s}_i^T\right) = \mathbf{Q}\mathbf{\Lambda}_2\mathbf{Q}^T \tag{71}$$

where $\mathbf{\Lambda}_1$ and $\mathbf{\Lambda}_2$ are diagonal matrices of dimension $L \times L$. Consider the permutation matrix $\mathbf{P}$ which partitions $\mathbf{Q}$ into $\tilde{\mathbf{Q}} \in \mathbb{R}^{L\times p}$ that contains the available eigenvectors of $\mathbf{R}_x$ with the corresponding eigenvalues in descending order and $\mathbf{Q}^* \in \mathbb{R}^{L\times(L-p)}$ that contains all used eigenvectors: $\mathbf{Q}\mathbf{P}^T = \begin{bmatrix} \tilde{\mathbf{Q}} & \mathbf{Q}^* \end{bmatrix}$. Let $\tilde{\mathbf{\Lambda}} \in \mathbb{R}^{p\times p}$ and $\mathbf{\Lambda}^* \in \mathbb{R}^{(L-p)\times(L-p)}$ be the diagonal matrices that contain the eigenvalues of the eigenvectors in $\tilde{\mathbf{Q}}$ and $\mathbf{Q}^*$, respectively. Then,

$$\mathbf{P}\mathbf{\Lambda}\mathbf{P}^T = \begin{bmatrix} \tilde{\mathbf{\Lambda}} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}^* \end{bmatrix},\ \mathbf{P}\mathbf{\Lambda}_1\mathbf{P}^T = \begin{bmatrix} \sigma_n^2\mathbf{I}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_1^* \end{bmatrix},\ \text{and}\ \mathbf{P}\mathbf{\Lambda}_2\mathbf{P}^T = \begin{bmatrix} \tilde{\mathbf{\Lambda}} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_2^* \end{bmatrix} \tag{72}$$

where $\mathbf{\Lambda}_1^*$ and $\mathbf{\Lambda}_2^*$ are diagonal matrices of dimension $(L-p)\times(L-p)$. Since $\mathbf{Q}$ is an orthonormal basis of $\mathbb{R}^L$ we can write $\mathbf{s}_{l+1}$ as

$$\mathbf{s}_{l+1} = \begin{bmatrix} \tilde{\mathbf{Q}} & \mathbf{Q}^* \end{bmatrix} \begin{bmatrix} \tilde{\boldsymbol{\beta}} \\ \boldsymbol{\beta}^* \end{bmatrix} = \mathbf{Q}\mathbf{P}^T \begin{bmatrix} \tilde{\boldsymbol{\beta}} \\ \boldsymbol{\beta}^* \end{bmatrix} \tag{73}$$

where $\tilde{\boldsymbol{\beta}} \in \mathbb{R}^p$ and $\boldsymbol{\beta}^* \in \mathbb{R}^{L-p}$. The constraint $\mathbf{s}_{l+1}^T\mathbf{s}_i = 0, i = 1, 2, \ldots, l$, is equivalent to $\boldsymbol{\beta}^* = \mathbf{0}$. Hence,

$$\mathbf{s}_{l+1} = \mathbf{Q}\mathbf{P}^T \begin{bmatrix} \tilde{\boldsymbol{\beta}} \\ \mathbf{0} \end{bmatrix}. \tag{74}$$

The unit norm requirement, $\|\mathbf{s}_{l+1}\| = 1$, implies that $\begin{bmatrix} \tilde{\boldsymbol{\beta}}^T & \mathbf{0}^T \end{bmatrix} \mathbf{P}\mathbf{Q}^T\mathbf{Q}\mathbf{P}^T \begin{bmatrix} \tilde{\boldsymbol{\beta}} \\ \mathbf{0} \end{bmatrix} = 1$ or $\left\|\tilde{\boldsymbol{\beta}}\right\| = 1$.

We want to find $\mathbf{s}_{l+1}^{\text{opt}} = \arg\max_{\mathbf{s}_{l+1}} \{\mathrm{SINR}_{\max}\}$ subject to $\|\mathbf{s}_{l+1}\| = 1$ and $\mathbf{s}_{l+1}^T\mathbf{s}_i = 0$, $i = 1, 2, \ldots, l$. Substituting $\mathbf{s}_{l+1}$ from (74) to the first term in (70) we obtain

$$\mathcal{D}_{l+1} - c_{l+1}^2 \mathbf{s}_{l+1}^T \mathbf{R}_x \mathbf{s}_{l+1} = \mathcal{D}_{l+1} - c_{l+1}^2 \tilde{\boldsymbol{\beta}}^T \tilde{\mathbf{\Lambda}} \tilde{\boldsymbol{\beta}}. \tag{75}$$

Using (71) and (74), the second term in (70) becomes

$$\left[\begin{array}{c}\tilde{\boldsymbol{\beta}}^T\mathbf{0}^T\end{array}\right]\left(\left(\mathbf{I}_L-c_{l+1}\left[\begin{array}{c}\tilde{\boldsymbol{\beta}}\\ \mathbf{0}\end{array}\right]\left[\begin{array}{c}\tilde{\boldsymbol{\beta}}^T\mathbf{0}^T\end{array}\right]\right)\mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^T\left(\mathbf{I}_L-c_{l+1}\left[\begin{array}{c}\tilde{\boldsymbol{\beta}}\\ \mathbf{0}\end{array}\right]\left[\begin{array}{c}\tilde{\boldsymbol{\beta}}^T\mathbf{0}^T\end{array}\right]\right)-\mathbf{P}(\boldsymbol{\Lambda}-\boldsymbol{\Lambda}_1-\boldsymbol{\Lambda}_2)\mathbf{P}^T\right)^{-1}\left[\begin{array}{c}\tilde{\boldsymbol{\beta}}\\ \mathbf{0}\end{array}\right] \quad (76)$$

which using (72) reduces to

$$\left[\begin{array}{cc}\tilde{\boldsymbol{\beta}}^T & \mathbf{0}^T\end{array}\right]\left[\begin{array}{cc}\left(\mathbf{I}_p-c_{l+1}\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\beta}}^T\right)\tilde{\boldsymbol{\Lambda}}\left(\mathbf{I}_p-c_{l+1}\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\beta}}^T\right)+\sigma_n^2\mathbf{I}_p & \mathbf{0}\\ \mathbf{0} & \tilde{\boldsymbol{\Lambda}}_1^*+\tilde{\boldsymbol{\Lambda}}_2^*\end{array}\right]^{-1}\left[\begin{array}{c}\tilde{\boldsymbol{\beta}}\\ \mathbf{0}\end{array}\right]. \quad (77)$$

We know that if a matrix $\mathbf{A}=\left[\begin{array}{cc}\mathbf{A}_{11} & \mathbf{A}_{12}\\ \mathbf{A}_{21} & \mathbf{A}_{22}\end{array}\right]$ is invertible and $\mathbf{A}_{11}^{-1}$, $\mathbf{A}_{22}^{-1}$ exist, then $\mathbf{A}^{-1}=$ $\left[\begin{array}{cc}\left(\mathbf{A}_{11}-\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\right)^{-1} & \left(\mathbf{A}_{11}-\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\right)^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\\ \left(\mathbf{A}_{22}-\mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\right)^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \left(\mathbf{A}_{22}-\mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\right)^{-1}\end{array}\right]$ [34]. Hence, (77) reduces further to

$$\tilde{\boldsymbol{\beta}}^T\left(\left(\mathbf{I}_p-c_{l+1}\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\beta}}^T\right)\tilde{\boldsymbol{\Lambda}}\left(\mathbf{I}_p-c_{l+1}\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\beta}}^T\right)+\sigma_n^2\mathbf{I}_p\right)^{-1}\tilde{\boldsymbol{\beta}}. \quad (78)$$

Using (75) and (78), the initial optimization problem can be written equivalently as

$$\tilde{\boldsymbol{\beta}}=\arg\max_{\tilde{\boldsymbol{\beta}}}\left\{\left(\mathcal{D}_{l+1}-c_{l+1}^2\tilde{\boldsymbol{\beta}}^T\tilde{\boldsymbol{\Lambda}}\tilde{\boldsymbol{\beta}}\right)\tilde{\boldsymbol{\beta}}^T\left(\left(\mathbf{I}_p-c_{l+1}\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\beta}}^T\right)\tilde{\boldsymbol{\Lambda}}\left(\mathbf{I}_p-c_{l+1}\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\beta}}^T\right)+\sigma_n^2\mathbf{I}_p\right)^{-1}\tilde{\boldsymbol{\beta}}\right\} \quad (79)$$

subject to $\left\|\tilde{\boldsymbol{\beta}}\right\|=1$. We will show that both terms in (79), $\mathcal{D}_{l+1}-c_{l+1}^2\tilde{\boldsymbol{\beta}}^T\tilde{\boldsymbol{\Lambda}}\tilde{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\beta}}^T\left(\left(\mathbf{I}_p-c_{l+1}\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\beta}}^T\right)\tilde{\boldsymbol{\Lambda}}\left(\mathbf{I}_p-c_{l+1}\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\beta}}^T\right)+\sigma_n^2\mathbf{I}_p\right)^{-1}\tilde{\boldsymbol{\beta}}$, are maximized by the same $\tilde{\boldsymbol{\beta}}$. For the first term we have:

$$\arg\max_{\tilde{\boldsymbol{\beta}}}\left\{\mathcal{D}_{l+1}-c_{l+1}^2\tilde{\boldsymbol{\beta}}^T\tilde{\boldsymbol{\Lambda}}\tilde{\boldsymbol{\beta}}\right\}=\arg\min_{\tilde{\boldsymbol{\beta}}}\left\{\tilde{\boldsymbol{\beta}}^T\tilde{\boldsymbol{\Lambda}}\tilde{\boldsymbol{\beta}}\right\}=\text{ minimum eigenvector of }\tilde{\boldsymbol{\Lambda}}. \quad (80)$$

Consider now the second term. By the matrix inversion lemma,

$$\begin{aligned}&\left(\left(\mathbf{I}_p-c_{l+1}\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\beta}}^T\right)\tilde{\boldsymbol{\Lambda}}\left(\mathbf{I}_p-c_{l+1}\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\beta}}^T\right)+\sigma_n^2\mathbf{I}_p\right)^{-1}\\&=\frac{1}{\sigma_n^2}\mathbf{I}_p-\frac{1}{\sigma_n^4}\mathbf{I}_p\left(\mathbf{I}_p-c_{l+1}\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\beta}}^T\right)\left(\tilde{\boldsymbol{\Lambda}}^{-1}+\frac{1}{\sigma_n^2}\left(\mathbf{I}_p-c_{l+1}\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\beta}}^T\right)^2\right)^{-1}\left(\mathbf{I}_p-c_{l+1}\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\beta}}^T\right).\end{aligned} \quad (81)$$

Using the matrix inversion lemma again,

$$\begin{aligned}&\left(\tilde{\boldsymbol{\Lambda}}^{-1}+\frac{1}{\sigma_n^2}\left(\mathbf{I}_p-c_{l+1}\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\beta}}^T\right)^2\right)^{-1}\\&=\left(\tilde{\boldsymbol{\Lambda}}^{-1}+\frac{1}{\sigma_n^2}\mathbf{I}_p\right)^{-1}-\frac{\left(c_{l+1}^2-2c_{l+1}\right)\left(\tilde{\boldsymbol{\Lambda}}^{-1}+\frac{1}{\sigma_n^2}\mathbf{I}_p\right)^{-1}\tilde{\boldsymbol{\beta}}\tilde{\boldsymbol{\beta}}^T\left(\tilde{\boldsymbol{\Lambda}}^{-1}+\frac{1}{\sigma_n^2}\mathbf{I}_p\right)^{-1}}{\sigma_n^2+\left(c_{l+1}^2-2c_{l+1}\right)\tilde{\boldsymbol{\beta}}^T\left(\tilde{\boldsymbol{\Lambda}}^{-1}+\frac{1}{\sigma_n^2}\mathbf{I}_p\right)^{-1}\tilde{\boldsymbol{\beta}}}.\end{aligned} \quad (82)$$

Combining (81) and (82) we have

$$\tilde{\boldsymbol{\beta}}^T \left(\left(\mathbf{I}_p - c_{l+1} \tilde{\boldsymbol{\beta}} \tilde{\boldsymbol{\beta}}^T \right) \tilde{\boldsymbol{\Lambda}} \left(\mathbf{I}_p - c_{l+1} \tilde{\boldsymbol{\beta}} \tilde{\boldsymbol{\beta}}^T \right) + \sigma_n^2 \mathbf{I}_p \right)^{-1} \tilde{\boldsymbol{\beta}} = \frac{1}{\sigma_n^2} \left(\frac{\sigma_n^2 - \tilde{\boldsymbol{\beta}}^T \left(\tilde{\boldsymbol{\Lambda}}^{-1} + \frac{1}{\sigma_n^2} \mathbf{I}_p \right)^{-1} \tilde{\boldsymbol{\beta}}}{\sigma_n^2 + \left(c_{l+1}^2 - 2c_{l+1} \right) \tilde{\boldsymbol{\beta}}^T \left(\tilde{\boldsymbol{\Lambda}}^{-1} + \frac{1}{\sigma_n^2} \mathbf{I}_p \right)^{-1} \tilde{\boldsymbol{\beta}}} \right). \tag{83}$$

Differentiation of the righthandside of (83) with respect to $\tilde{\boldsymbol{\beta}}^T \left(\tilde{\boldsymbol{\Lambda}}^{-1} + \frac{1}{\sigma_n^2} \mathbf{I}_p \right)^{-1} \tilde{\boldsymbol{\beta}}$ gives $\frac{1}{\sigma_n^2} \frac{-\sigma_n^2 (1 + c_{l+1}(c_{l+1}-2))}{\left(\sigma_n^2 + c_{l+1}(c_{l+1}-2) \tilde{\boldsymbol{\beta}}^T \left(\tilde{\boldsymbol{\Lambda}}^{-1} + \frac{1}{\sigma_n^2} \mathbf{I}_p \right)^{-1} \tilde{\boldsymbol{\beta}} \right)^2} \leq 0 \quad \forall c_{l+1} \in \mathbb{R}$. Hence, $\frac{1}{\sigma_n^2} \frac{\sigma_n^2 - \tilde{\boldsymbol{\beta}}^T \left(\tilde{\boldsymbol{\Lambda}}^{-1} + \frac{1}{\sigma_n^2} \mathbf{I}_p \right)^{-1} \tilde{\boldsymbol{\beta}}}{\left(\sigma_n^2 + c_{l+1}(c_{l+1}-2) \tilde{\boldsymbol{\beta}}^T \left(\tilde{\boldsymbol{\Lambda}}^{-1} + \frac{1}{\sigma_n^2} \mathbf{I}_p \right)^{-1} \tilde{\boldsymbol{\beta}} \right)}$ is a decreasing function of $\tilde{\boldsymbol{\beta}}^T \left(\tilde{\boldsymbol{\Lambda}}^{-1} + \frac{1}{\sigma_n^2} \mathbf{I}_p \right)^{-1} \tilde{\boldsymbol{\beta}}$. However, $\frac{\tilde{\lambda}_{min}}{1 + \frac{\tilde{\lambda}_{min}}{\sigma_n^2}} \leq \tilde{\boldsymbol{\beta}}^T \left(\tilde{\boldsymbol{\Lambda}}^{-1} + \frac{1}{\sigma_n^2} \mathbf{I}_p \right)^{-1} \tilde{\boldsymbol{\beta}} \leq \frac{\tilde{\lambda}_{max}}{1 + \frac{\tilde{\lambda}_{max}}{\sigma_n^2}}$ where $\tilde{\lambda}_{min}$ and $\tilde{\lambda}_{max}$ are the minimum and maximum eigenvalues in $\tilde{\boldsymbol{\Lambda}}$. Therefore,

$$\arg\max_{\tilde{\boldsymbol{\beta}}} \left\{ \tilde{\boldsymbol{\beta}}^T \left(\left(\mathbf{I}_p - c_{l+1} \tilde{\boldsymbol{\beta}} \tilde{\boldsymbol{\beta}}^T \right) \tilde{\boldsymbol{\Lambda}} \left(\mathbf{I}_p - c_{l+1} \tilde{\boldsymbol{\beta}} \tilde{\boldsymbol{\beta}}^T \right) + \sigma_n^2 \mathbf{I}_p \right)^{-1} \tilde{\boldsymbol{\beta}} \right\} = \text{minimum eigenvector of } \tilde{\boldsymbol{\Lambda}}. \tag{84}$$

We conclude (cf. (79), (80), (84), and (74)) that the SINR expression in (70) is maximized when $\mathbf{s}_{l+1}$ is the minimum available eigenvector of $\mathbf{R}_x$ for any c_{l+1}. Hence,

$$\begin{aligned} \max_{c_{l+1}, \mathbf{s}_{l+1}} \{\text{SINR}_{\max}\} &= \max_{c_{l+1}} \left\{ \max_{\tilde{\boldsymbol{\beta}}} \left\{ \left(\mathcal{D}_{l+1} - c_{l+1}^2 \tilde{\boldsymbol{\beta}}^T \tilde{\boldsymbol{\Lambda}} \tilde{\boldsymbol{\beta}} \right) \tilde{\boldsymbol{\beta}}^T \left(\left(\mathbf{I}_p - c_{l+1} \tilde{\boldsymbol{\beta}} \tilde{\boldsymbol{\beta}}^T \right) \tilde{\boldsymbol{\Lambda}} \left(\mathbf{I}_p - c_{l+1} \tilde{\boldsymbol{\beta}} \tilde{\boldsymbol{\beta}}^T \right) + \sigma_n^2 \mathbf{I}_p \right)^{-1} \tilde{\boldsymbol{\beta}} \right\} \right\} \\ &= \max_{c_{l+1}} \left\{ \frac{\mathcal{D}_{l+1} - c_{l+1}^2 \lambda_j}{\sigma_n^2 + \lambda_j + c_{l+1}(c_{l+1}-2)\lambda_j} \right\} \end{aligned} \tag{85}$$

where λ_j is the minimum available eigenvalue of $\mathbf{R}_x$ (equivalently λ_j is the bottom element of $\tilde{\boldsymbol{\Lambda}}, \lambda_j = \tilde{\lambda}_{min}$). The optimum c_{l+1} value can be computed by setting the derivative of the last expression in (85) equal to zero. The latter gives two candidate values for c_{l+1}. We select the value that maximizes (85), which is the one in (54). □

References

[1] F. Hartung and M. Kutter, "Multimedia watermarking techniques," *Proc. of IEEE*, vol. 87, pp. 1079-1107, July 1999.

[2] G. C. Langelaar, I. Setyawan, and R. L. Lagendijk, "Watermarking digital image and video data: A state-of-the-art overview," *IEEE Signal Proc. Magazine*, vol. 17, pp. 20-46, Sept. 2000.

[3] L. Marvel and C. G. Boncelet, "Spread spectrum image steganography," *IEEE Trans. Image Proc.*, vol. 8, pp. 1075-1083, Aug. 1999.

[4] M. Kutter and S. Winkler, "A vision-based masking model for spread-spectrum image watermarking," *IEEE Trans. Image Proc.*, vol. 11, pp. 16-25, Jan. 2002.

[5] J. R. Smith and B. O. Comiskey, "Modulation and information hiding in images," *Springer-Verlag Lecture Notes in Computer Science*, vol. 1174, pp. 207-226, 1996.

[6] M. Wu and B. Liu, "Data hiding in binary image for authentication and annotation," *IEEE Trans. Multimedia*, vol. 6, pp. 528-538, Aug. 2004.

[7] I. J. Cox, J. Kilian, F. T. Leighton, and T. Shannon, "Secure spread spectrum watermarking for multimedia," *IEEE Trans. Image Proc.*, vol. 6, pp. 1673-1687, Dec. 1997.

[8] M. Barni, F. Bartolini, A. De Rosa, and A. Piva, "Optimum decoding and detection of multiplicative watermarks," *IEEE Trans. Signal Proc.*, vol. 51, pp. 1118-1123, Apr. 2003.

[9] M. Barni, F. Bartolini, A. De Rosa, and A. Piva, "A new decoder for the optimum recovery of nonadditive watermarks," *IEEE Trans. Image Proc.*, vol. 10, pp. 755-766, May 2001.

[10] C. Qiang and T. S. Huang, "Robust optimum detection of transform domain multiplicative watermarks," *IEEE Trans. Signal Proc.*, vol. 51, pp. 906-924, Apr. 2003.

[11] M. Barni, F. Bartolini, A. De Rosa, and A. Piva, "Capacity of full frame DCT image watermarks," *IEEE Trans. Image Proc.*, vol. 9, pp. 1450-1455, Aug. 2000.

[12] J. Hernandez, M. Amado, and F. Perez-Gonzalez, "DCT-domain watermarking techniques for still images: Detector performance analysis and a new structure," *IEEE Trans. Image Proc.*, vol. 9, pp. 55-68, Jan. 2000.

[13] C. Qiang and T. S. Huang, "An additive approach to transform-domain information hiding and optimum detection structure," *IEEE Trans. Multimedia*, vol. 3, pp. 273-284, Sept. 2001.

[14] S. Pereira, S. Voloshynovskiy, and T. Pun, "Optimized wavelet domain watermark embedding strategy using linear programming," in *Proc. of SPIE, Wavelet Applications Conf.*, Orlando, FL, April 2000, vol. 4056, pp. 490-498.

[15] P. Moulin and A. Ivanović, "The zero-rate spread-spectrum watermarking game," *IEEE Trans. Signal Proc.*, vol. 51, pp. 1098-1117, Apr. 2003.

[16] X. G. Xia, C. G. Boncelet, and G. R. Arce, "A multiresolution watermark for digital images," in *Proc. IEEE Int. Conf. Image Proc.*, vol. 1, Nov. 1998, pp. 548-551.

[17] C. Fei, D. Kundur, and R. H. Kwong, "Analysis and design of watermarking algorithms for improved resistance to compression," *IEEE Trans. Image Proc.*, vol. 13, pp. 126-144, Feb. 2004.

[18] H. S. Malvar and D. A. Florêncio, "Improved spread spectrum: A new modulation technique for robust watermarking," *IEEE Trans. Signal Proc.*, vol. 51, pp. 898-905, Apr. 2003.

[19] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. Inform. Theory*, vol. IT-29, pp. 439-441, May 1983.

[20] B. Chen and G. Wornell, "Quantization index modulation: A class of provably good methods for digital watermarking and information embedding," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1423-1443, May 2001.

[21] P. Moulin and J. A. O'Sullivan, "Information-theoretic analysis of information hiding," *IEEE Trans. Inform. Theory*, vol. 49, pp. 563-593, March 2003.

[22] H. L. Van Trees, *Detection Estimation and Modulation Theory, Part I.* New York: Wiley, 2001.

[23] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* New York: Wiley, 1991.

[24] C. Rose, S. Ulukus, R. D. Yates "Wireless systems and interference avoidance," *IEEE Trans. Wireless Commun.*, vol. 1, pp. 415-428, July 2002.

[25] C. Rose, "CDMA codeword optimization: Interference avoidance and convergence via class warfare," *IEEE Trans. Inform. Theory*, vol. 47, pp. 2368-2382, Sept. 2001.

[26] P. Anigstein and V. Anantharam, "Ensuring convergence of the MMSE iteration for interference avoidance to the global optimum," *IEEE Trans. Inform. Theory*, vol. 49, pp. 873-885, Apr. 2003.

[27] M. Rupf and J. L. Massey, "Optimum sequence multisets for syncronous code-division-multiple-access channels," *IEEE Trans. Inform. Theory*, vol. IT-40, pp. 1261-1266, July 1994.

[28] P. Viswanath and V. Anantharam, "Optimal sequences for CDMA under colored noise: A Schur-Saddle function property," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1295-1318, June 2002.

[29] G. N. Karystinos and D. A. Pados, "Code division multiplexing performance of minimum total-squared-correlation binary signature sets," in *Proc. IEEE Globecom*, Comm. Theory Symposium, vol. 4, San Fransisco, CA, Dec. 2003, pp. 1862-1866.

[30] G. N. Karystinos and D. A. Pados, "The maximum squared correlation, sum capacity, and total asymptotic efficiency of minimum total-squared-correlation binary signature sets," *IEEE Trans. Inform. Theory*, to appear Jan. 2005.

[31] P. Viswanath, D. N. C. Tse, V. Anantharam, "Asymptotically optimal waterfilling in multiple antenna multiple access channels," *IEEE Trans. Inform. Theory*, vol. 47, pp. 241-267, Jan. 2001.

[32] P. Moulin and M. K. Mihçak, "A framework for evaluating the data-hiding capacity of image sources," *IEEE Trans. Image Proc.*, vol. 11, pp. 1029-1042, Sept. 2002.

[33] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra.* Philadelphia, PA: SIAM, 2000.

[34] D. G. Manolakis, V. K. Ingle, and S. M. Kogon, *Statistical and Adaptive Signal Processing.* New York: McGraw-Hill, 2000.